

Substitute Sequence Listing_USSN 10587032_PP019482.007
 SUBSTITUTE SEQUENCE LISTING

<110> CHIRON CORPORATION
 HARDY, Stephen F
 DONNELLY, III, John J
 ZUR MEGEDE, Jan T

<120> VECTORS FOR EXPRESSION OF HML-2 POLYPEPTIDES

<130> PP19482.0007

<140> 10/587,032
 <141> 2006-07-24

<150> PCT/US03/18666
 <151> 2003-06-13

<150> 60/388831
 <151> 2002-06-16

<150> 60/472189
 <151> 2003-05-20

<160> 83

<170> PatentIn, version 3.5

<210> 1
 <211> 1998
 <212> DNA
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 1

atggggcaaa	ctaaaagtaa	aattaaaagt	aaatatgcct	cttatctcag	ctttatttaa	60
attcttttaa	aaagaggggg	agttaaagta	tctacaaaaa	atctaataca	gctattttcaa	120
ataatagaac	aattttgccc	atggttttcca	gaacaaggaa	cttttagatct	aaaagattgg	180
aaaagaattg	gtaaggaact	aaaacaagca	ggttaggaagg	gtaatatcat	tccacttaca	240
gtatgggaatg	attggggccat	tattaaagca	gctttagaac	catttcaaac	agaagaagat	300
agcgtttcag	tttctgatgc	ccctggaagc	tgtataatag	attgtaatga	aaacacaagg	360
aaaaaatccc	agaaagaaac	ggaagggttta	cattgcgaat	atgtagcaga	gccggtaatg	420
gctcagtcac	cgcaaaatgt	tgactataat	caattacagg	aggtagatata	tcctgaaacg	480
ttaaaattag	aaggaaaagg	tccagaatta	gtggggccat	cagagtctaa	accacgaggg	540
acaagtcctc	ttccagcagg	tcaggtgcct	gtaacattac	aacctcaaaa	gcagggttaa	600
gaaaataaga	cccaaccgcc	agtagcctat	caatactggc	ctccggctga	acttcagtat	660
cggccacccc	cagaaagtca	gtatggatat	ccaggaatgc	ccccagcacc	acagggcagg	720
gcgccatacc	ctcagccgcc	cactaggaga	cttaatccta	cggcaccacc	tagtagacag	780
ggtagtaaat	tacatgaaat	tattgataaa	tcaagaaagg	aaggagatac	tgaggcatgg	840
caattcccag	taacgttaga	accgatgcc	cctggagaag	gagcccaaga	gggagagcct	900
cccacagttg	aggccagata	caagtctttt	tcgataaaaa	agctaaaaga	tatgaaagag	960
ggagtaaaac	agtatggacc	caactcccct	tatatgagga	cattattaga	ttccattgct	1020
catggacata	gactcattcc	ttatgattgg	gagattctgg	caaaatcgct	tctctcacc	1080
tctcaatttt	tacaatttaa	gacttggtgg	attgatgggg	tacaagaaca	ggtccgaaga	1140
aatagggctg	ccaatcctcc	agttaacata	gatgcagatc	aactattagg	aatagggtcaa	1200
aattggagta	ctattagtca	acaagcatta	atgcaaaatg	aggccattga	gcaagttaga	1260
gctatctgcc	ttagagcctg	ggaaaaaatc	caagaccagg	gaagtacctg	cccctcattt	1320
aatacagtaa	gacaaggttc	aaaagagccc	tatcctgatt	ttgtggcaag	gctccaagat	1380
gttgctcaaa	agtcaattgc	tgatgaaaaa	gcccgtgaag	tcatagtgga	gttgatggca	1440
tatgaaaacg	ccaatcctga	gtgtcaatca	gccattaagc	cattaaaagg	aaaggttcct	1500
gcaggatcag	atgtaatctc	agaatatgta	aaagcctgtg	atggaatcgg	aggagctatg	1560
cataaagcta	tgcttatggc	tcaagcaata	acaggagttg	ttttaggagg	acaagttaga	1620
acatttgtaa	gaaaatgtta	taattgtggt	caaattggtc	acttaaaaaa	gaattgccca	1680
gtcttaataa	aacagaatat	aactattcaa	gcaactacaa	caggtagaga	gccacctgac	1740

Substitute Sequence Listing_USSN 10587032_PP019482.007

ttatgtccaa	gatgtaaaaa	aggaaaacat	tgggctagtc	aatgtcgttc	taaatttgat	1800
aaaaatgggc	aaccattgtc	gggaaacgag	caaagggggc	agcctcaggc	cccacaacaa	1860
actggggcat	tcccaattca	gccatttgtt	cctcaggggt	ttcagggaca	acaacccccca	1920
ctgtcccaag	tgtttcaggg	aataagccag	ttaccacaat	acaacaattg	tcccccgcca	1980
caagcggcag	tgcagcag					1998

<210> 2
 <211> 2001
 <212> DNA
 <213> Human endogenous retrovirus, K family (HERV-K)

<400>	2					
atggggcaaa	ctaaaagtaa	aattaaagt	aaatatgcct	cttatctcag	ctttattaaa	60
attcttttaa	aaagaggggg	agttaaagta	tctacaaaaa	atctaataca	gctattttcaa	120
ataatagaac	aattttgccc	atggttttcca	gaacaaggaa	ctttagatct	aaaagatttg	180
aaaagaattg	gtaaggaact	aaaacaagca	ggtaggaagg	gtaatatcat	tccacttaca	240
gtatggaatg	attgggccat	tattaaagca	gctttagaac	catttcaaac	agaagaagat	300
agcgtttcag	tttctgatgc	ccctggaagc	tgtataatag	attgtaatga	aaacacaagg	360
aaaaaatccc	agaaagaaac	ggaaggttta	cattgcgaat	atgtagcaga	gccggtaaatg	420
gctcagtcaa	cgcaaaatgt	tgactataat	caattacagg	aggtgatata	tcctgaaacg	480
ttaaaattag	aaggaaaagg	tccagaatta	gtggggccat	cagagtctaa	accacgaggg	540
acaagtcctc	ttccagcagg	tcaggtgcct	gtaacattac	aacctcaaaa	gcagggttaa	600
gaaaataaga	cccaaccgcc	agtagcctat	caatactggc	ctccggctga	acttcagtat	660
cggccacccc	cagaaagtca	gtatggatat	ccaggaatgc	ccccagcacc	acagggcagg	720
gcgccatacc	ctcagccgcc	cactaggaga	cttaatccta	cggcaccacc	tagtagacag	780
ggtagtaaat	tacatgaaat	tattgataaa	tcaagaaagg	aaggagatac	tgaggcatgg	840
caattcccaag	taacgttaga	accgatgcca	cctggagaag	gagcccaaga	gggagagcct	900
cccacagttg	aggccagata	caagtctttt	tcgataaaaa	agctgaaaga	tatgaaagag	960
ggagtaaaac	agtatggacc	caactccctt	tatatgagga	cattattaga	ttccattgct	1020
catggacata	gactcattcc	ttatgattgg	gagattctgg	caaaatcgtc	tctctcacc	1080
tctcaatttt	tacaatttaa	gacttggtgg	attgatgggg	tacaagaaca	ggtccgaaga	1140
aatagggtcg	ccaatcctcc	agttaacata	gatgcagatc	aactattagg	aataggtcaa	1200
aattggagta	ctattagtca	acaagcatta	atgcaaaatg	aggccattga	gcaagttaga	1260
gctatctgcc	ttagagcctg	ggaaaaaatc	caagaccag	gaagtacctg	ccccctcattt	1320
aatacagtaa	gacaagggtc	aaaagagccc	tatcctgatt	ttgtggcaag	gctccaagat	1380
gttgctcaaa	agtcaattgc	tgatgaaaaa	gcccgtaagg	tcatagtggg	gttgatggca	1440
tatgaaaacg	ccaatcctga	gtgtcaatca	gccattaagc	cattaaaagg	aaagggttctt	1500
gcaggatcag	atgtaatctc	agaatatgta	aaagcctgtg	atggaatcgg	aggagctatg	1560
tataaagcta	tgcttatggc	tcaagcaata	acaggagtgt	ttttaggagg	acaagttaga	1620
acatttgga	gaaaatgtta	taattgtggt	caaattgggtc	acttaaaaaa	gaattgcccc	1680
gtcttaaaata	aacagaatat	aactattcaa	gcaactacaa	caggtagaga	gccacctgac	1740
ttatgtccaa	gatgtaaaaa	aggaaaacat	tgggctagtc	aatgtcgttc	taaatttgat	1800
aaaaatgggc	aaccattgtc	gggaaacgag	caaagggggc	agcctcaggc	cccacaacaa	1860
actggggcat	tcccaattca	gccatttgtt	cctcaggggt	ttcagggaca	acaacccccca	1920
ctgtcccaag	tgtttcaggg	aataagccag	ttaccacaat	acaacaattg	tcccccgcca	1980
caagcggcag	tgcagcagta	g				2001

<210> 3
 <211> 2004
 <212> DNA
 <213> Human endogenous retrovirus, K family (HERV-K)

<400>	3					
atggggcaaa	ctaaaagtaa	aactaaaagt	aaatatgcct	cttatctcag	ctttattaaa	60
attcttttaa	aaagaggggg	agtttagagta	tctacaaaaa	atctaataca	gctattttcaa	120
ataatagaac	aattttgccc	atggttttcca	gaacaaggaa	ctttagatct	aaaagatttg	180
aaaagaattg	gcgaggaact	aaaacaagca	ggtagaaagg	gtaatatcat	tccacttaca	240
gtatggaatg	attgggccat	tattaaagca	gctttagaac	catttcaaac	aaaagaagat	300
agcgtttcag	tttctgatgc	ccctggaagc	tgtgtaatag	attgtaatga	aaagacaggg	360
agaaaatccc	agaaagaaac	agaaagttaa	cattgcgaat	atgtaacaga	gccagtaaatg	420
gctcagtcaa	cgcaaaatgt	tgactataat	caattacagg	gggtgatata	tcctgaaacg	480
ttaaaattag	aaggaaaagg	tccagaatta	gtggggccat	cagagtctaa	accacgaggg	540
ccaagtcctc	ttccagcagg	tcaggtgccc	gtaacattac	aacctcaaac	gcagggttaa	600

Substitute Sequence Listing_USSN 10587032_PP019482.007

gaaaataaga	cccaaccgcc	agtagcttat	caatactggc	cgccggctga	acttcagtat	660
ctgccacccc	cagaaagtca	gtatggatat	ccaggaatgc	ccccagcact	acagggcagg	720
gcgccatatt	ctcagccgcc	cactgtgaga	cttaatccta	cagcatcacg	tagtggacaa	780
ggtggtacac	tgcacgcagt	cattgatgaa	gccagaaaac	agggagatct	tgaggcatgg	840
cggttcctgg	taattttaca	actggtacag	gccggggaag	agactcaagt	aggagcgctt	900
gcccagagctg	agactagatg	tgaacctttc	accatgaaaa	tgttaaaaga	tataaaggaa	960
ggagttaaac	aatatggatc	caactcccct	tatataagaa	cattattaga	ttccattgct	1020
catggaataa	gacttactcc	ttatgactgg	gaaagtttgg	ccaaatcttc	cctttcatcc	1080
tctcagtatc	tacagtttaa	aacctggtgg	attgatggag	tacaagaaca	ggtacgaaaa	1140
aatcaggcta	ctaagcccac	tgtaatatata	gacgcagacc	aattgttagg	aacagggtcca	1200
aattggagca	ccattaacca	acaatcagtg	atgcagaatg	aggctattga	acaagtaagg	1260
gctatttgcc	tcagggcctg	gggaaaaatt	caggacccag	gaacagcttt	ccctattaat	1320
tcaattagac	aaggctctaa	agagccatat	cctgactttg	tggaagatt	acaagatgct	1380
gctcaaaagt	ctattacaga	tgacaatgcc	cgaaaagtta	ttgtagaatt	aatggcctat	1440
gaaaatgcaa	atccagaatg	tcagtcggcc	ataaagccat	taaaaggaaa	agttccagca	1500
ggagttgatg	taattacaga	atatgtgaag	gcttgtgatg	ggattggagg	agctatgcat	1560
aaggcaatgc	taatggctca	agcaatgagg	gggctcactc	taggaggaca	agttagaaca	1620
tttgggaaaa	aatgttataa	ttgtgggtcaa	atcggtcatc	tgaaaaggag	ttgcccagtc	1680
ttaaataaac	agaatataat	aaatcaagct	attacagcaa	aaaataaaaa	gccatctggc	1740
ctgtgtccaa	aatgtggaaa	aggaaaacat	tgggccaatc	aatgtcattc	taaatttgat	1800
aaagatgggc	aaccattgtc	gggaaacagg	aagaggggcc	agcctcaggc	cccccaacaa	1860
actggggcat	tcccagttca	actgtttgtt	cctcagggtt	ttcaaggaca	acaaccctta	1920
cagaaaatac	caccacttca	gggagtcagc	caattacaac	aatccaacag	ctgtcccgcg	1980
ccacagcagg	cagcgccaca	gtag				2004

<210> 4
 <211> 852
 <212> DNA
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 4						
atggggcaaa	ctaaaagtaa	aattaaaagt	aaatatgcct	cttatctcag	ctttattaaa	60
attcttttaa	aaagaggggg	agttaaagta	tctacaaaaa	atctaataca	gctattttcaa	120
ataatagaac	aattttgccc	atggtttcca	gaacaaggaa	cttcagatct	aaaagattgg	180
aaaagaattg	gtaaggaact	aaaacaagca	ggtaggaagg	gtaatatcat	tccacttaca	240
gtatggaatg	attgggccat	tattaaagca	gctttagaac	catttcaaac	agaagaagat	300
agcatttcag	tttctgatgc	ccctggaagc	tgtttaatat	attgtaatga	aaacacaagg	360
aaaaaatccc	agaaagaaac	cgaaagttaa	cattgcgaat	atgtagcaga	gccggtaatg	420
gctcagtcac	cgcaaaatgt	tgactataat	caattacagg	aggatgata	tcctgaaacg	480
ttaaaattag	aaggaaaagg	tccagaatta	atggggccat	cagagtctaa	accacgaggc	540
acaagtcctc	ttccagcagg	tcaggtgctc	gtaagattac	aacctcaaaa	gcagggttaa	600
gaaaataaga	cccaaccgca	agtagcctat	caatactgcc	gctggctgaa	cttcagtatc	660
ggccaccccc	agaaagtcag	tatggatatc	caggaatgcc	cccagcacca	cagggcaggg	720
cgccatacca	tcagccgccc	actaggagac	ttaatcctat	ggcaccacct	agtagacagg	780
gtagtgaatt	acatgaaatt	attgataaat	caagaaagga	aggagatact	gaggcatggc	840
aattcccagt	aa					852

<210> 5
 <211> 666
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 5						
Met Gly Gln Thr	Lys Ser Lys Ile	Lys Ser Lys Tyr	Ala Ser Tyr	Leu		
1	5	10	15			
Ser Phe Ile	Lys Ile Leu Leu	Lys Arg Gly Gly	Val Lys Val	Ser Thr		
	20	25	30			
Lys Asn Leu	Ile Lys Leu Phe	Gln Ile Ile Glu	Gln Phe Cys	Pro Trp		
	35	40	45			
Phe Pro Glu	Gln Gly Thr Leu	Asp Leu Lys Asp	Trp Lys Arg	Ile Gly		

Substitute Sequence Listing_USSN 10587032_PP019482.007

50

55

60

Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65 70 75 80
Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
85 90 95
Thr Glu Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Ile
100 105 110
Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
115 120 125
Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
130 135 140
Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145 150 155 160
Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
165 170 175
Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
180 185 190
Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
195 200 205
Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
210 215 220
Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
225 230 235 240
Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
245 250 255
Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
260 265 270
Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
275 280 285
Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
290 295 300
Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
305 310 315 320
Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
325 330 335
Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
340 345 350
Leu Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
355 360 365
Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
370 375 380
Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln

Substitute Sequence Listing_USSN 10587032_PP019482.007

385 390 395 400
 Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
 405 410 415
 Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Glu Lys Ile Gln Asp
 420 425 430
 Pro Gly Ser Thr Cys Pro Ser Phe Asn Thr Val Arg Gln Gly Ser Lys
 435 440 445
 Glu Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Val Ala Gln Lys
 450 455 460
 Ser Ile Ala Asp Glu Lys Ala Arg Lys Val Ile Val Glu Leu Met Ala
 465 470 475 480
 Tyr Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys
 485 490 495
 Gly Lys Val Pro Ala Gly Ser Asp Val Ile Ser Glu Tyr Val Lys Ala
 500 505 510
 Cys Asp Gly Ile Gly Gly Ala Met Tyr Lys Ala Met Leu Met Ala Gln
 515 520 525
 Ala Ile Thr Gly Val Val Leu Gly Gly Gln Val Arg Thr Phe Gly Arg
 530 535 540
 Lys Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Lys Asn Cys Pro
 545 550 555 560
 Val Leu Asn Lys Gln Asn Ile Thr Ile Gln Ala Thr Thr Thr Gly Arg
 565 570 575
 Glu Pro Pro Asp Leu Cys Pro Arg Cys Lys Lys Gly Lys His Trp Ala
 580 585 590
 Ser Gln Cys Arg Ser Lys Phe Asp Lys Asn Gly Gln Pro Leu Ser Gly
 595 600 605
 Asn Glu Gln Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
 610 615 620
 Pro Ile Gln Pro Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Pro
 625 630 635 640
 Leu Ser Gln Val Phe Gln Gly Ile Ser Gln Leu Pro Gln Tyr Asn Asn
 645 650 655
 Cys Pro Pro Pro Gln Ala Ala Val Gln Gln
 660 665

<210>

6

<211>

667

<212>

PRT

<213>

Human endogenous retrovirus, K family (HERV-K)

<400>

6

Met Gly Gln Thr Lys Ser Lys Thr Lys Ser Lys Tyr Ala Ser Tyr Leu
 1 5 10 15

Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Arg Val Ser Thr

Substitute Sequence Listing_USSN 10587032_PP019482.007

Lys	Asn	Leu	Ile	Lys	Leu	Phe	Gln	Ile	Ile	Glu	Gln	Phe	Cys	Pro	Trp
		35					40					45			
Phe	Pro	Glu	Gln	Gly	Thr	Leu	Asp	Leu	Lys	Asp	Trp	Lys	Arg	Ile	Gly
	50					55					60				
Glu	Glu	Leu	Lys	Gln	Ala	Gly	Arg	Lys	Gly	Asn	Ile	Ile	Pro	Leu	Thr
65					70					75					80
Val	Trp	Asn	Asp	Trp	Ala	Ile	Ile	Lys	Ala	Ala	Leu	Glu	Pro	Phe	Gln
			85						90					95	
Thr	Lys	Glu	Asp	Ser	Val	Ser	Val	Ser	Asp	Ala	Pro	Gly	Ser	Cys	Val
			100					105					110		
Ile	Asp	Cys	Asn	Glu	Lys	Thr	Gly	Arg	Lys	Ser	Gln	Lys	Glu	Thr	Glu
		115					120					125			
Ser	Leu	His	Cys	Glu	Tyr	Val	Thr	Glu	Pro	Val	Met	Ala	Gln	Ser	Thr
	130					135					140				
Gln	Asn	Val	Asp	Tyr	Asn	Gln	Leu	Gln	Gly	Val	Ile	Tyr	Pro	Glu	Thr
145					150					155					160
Leu	Lys	Leu	Glu	Gly	Lys	Gly	Pro	Glu	Leu	Val	Gly	Pro	Ser	Glu	Ser
				165					170					175	
Lys	Pro	Arg	Gly	Pro	Ser	Pro	Leu	Pro	Ala	Gly	Gln	Val	Pro	Val	Thr
			180					185					190		
Leu	Gln	Pro	Gln	Thr	Gln	Val	Lys	Glu	Asn	Lys	Thr	Gln	Pro	Pro	Val
		195					200					205			
Ala	Tyr	Gln	Tyr	Trp	Pro	Pro	Ala	Glu	Leu	Gln	Tyr	Leu	Pro	Pro	Pro
	210					215					220				
Glu	Ser	Gln	Tyr	Gly	Tyr	Pro	Gly	Met	Pro	Pro	Ala	Leu	Gln	Gly	Arg
225					230					235					240
Ala	Pro	Tyr	Pro	Gln	Pro	Pro	Thr	Val	Arg	Leu	Asn	Pro	Thr	Ala	Ser
				245					250					255	
Arg	Ser	Gly	Gln	Gly	Gly	Thr	Leu	His	Ala	Val	Ile	Asp	Glu	Ala	Arg
			260					265					270		
Lys	Gln	Gly	Asp	Leu	Glu	Ala	Trp	Arg	Phe	Leu	Val	Ile	Leu	Gln	Leu
		275					280					285			
Val	Gln	Ala	Gly	Glu	Glu	Thr	Gln	Val	Gly	Ala	Pro	Ala	Arg	Ala	Glu
	290					295					300				
Thr	Arg	Cys	Glu	Pro	Phe	Thr	Met	Lys	Met	Leu	Lys	Asp	Ile	Lys	Glu
305					310					315					320
Gly	Val	Lys	Gln	Tyr	Gly	Ser	Asn	Ser	Pro	Tyr	Ile	Arg	Thr	Leu	Leu
				325					330					335	
Asp	Ser	Ile	Ala	His	Gly	Asn	Arg	Leu	Thr	Pro	Tyr	Asp	Trp	Glu	Ser
			340					345					350		
Leu	Ala	Lys	Ser	Ser	Leu	Ser	Ser	Ser	Gln	Tyr	Leu	Gln	Phe	Lys	Thr

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Lys Asn Gln Ala Thr
370 375 380

Lys Pro Thr Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Thr Gly Pro
385 390 395 400

Asn Trp Ser Thr Ile Asn Gln Gln Ser Val Met Gln Asn Glu Ala Ile
405 410 415

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Gly Lys Ile Gln Asp
420 425 430

Pro Gly Thr Ala Phe Pro Ile Asn Ser Ile Arg Gln Gly Ser Lys Glu
435 440 445

Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Ala Ala Gln Lys Ser
450 455 460

Ile Thr Asp Asp Asn Ala Arg Lys Val Ile Val Glu Leu Met Ala Tyr
465 470 475 480

Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys Gly
485 490 495

Lys Val Pro Ala Gly Val Asp Val Ile Thr Glu Tyr Val Lys Ala Cys
500 505 510

Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln Ala
515 520 525

Met Arg Gly Leu Thr Leu Gly Gly Gln Val Arg Thr Phe Gly Lys Lys
530 535 540

Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Arg Ser Cys Pro Val
545 550 555 560

Leu Asn Lys Gln Asn Ile Ile Asn Gln Ala Ile Thr Ala Lys Asn Lys
565 570 575

Lys Pro Ser Gly Leu Cys Pro Lys Cys Gly Lys Gly Lys His Trp Ala
580 585 590

Asn Gln Cys His Ser Lys Phe Asp Lys Asp Gly Gln Pro Leu Ser Gly
595 600 605

Asn Arg Lys Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
610 615 620

Pro Val Gln Leu Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Leu
625 630 635 640

Gln Lys Ile Pro Pro Leu Gln Gly Val Ser Gln Leu Gln Gln Ser Asn
645 650 655

Ser Cys Pro Ala Pro Gln Gln Ala Ala Pro Gln
660 665

Substitute Sequence Listing_USSN 10587032_PP019482.007

<400> 7
Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
1 5 10 15
Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
20 25 30
Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
35 40 45
Phe Pro Glu Gln Gly Thr Ser Asp Leu Lys Asp Trp Lys Arg Ile Gly
50 55 60
Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65 70 75 80
Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
85 90 95
Thr Glu Glu Asp Ser Ile Ser Val Ser Asp Ala Pro Gly Ser Cys Leu
100 105 110
Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
115 120 125
Ser Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
130 135 140
Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145 150 155 160
Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Met Gly Pro Ser Glu Ser
165 170 175
Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Leu Val Arg
180 185 190
Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Gln Val
195 200 205
Ala Tyr Gln Tyr Cys Arg Trp Leu Asn Phe Ser Ile Gly His Pro Gln
210 215 220
Lys Val Ser Met Asp Ile Gln Glu Cys Pro Gln His His Arg Ala Gly
225 230 235 240
Arg His Thr Ile Ser Arg Pro Leu Gly Asp Leu Ile Leu Trp His His
245 250 255
Leu Val Asp Arg Val Val Asn Tyr Met Lys Leu Leu Ile Asn Gln Glu
260 265 270
Arg Lys Glu Ile Leu Arg His Gly Asn Ser Gln
275 280

<210> 8

<211> 434

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 8

Met Pro Pro Ala Pro Gln Gly Arg Ala Pro Tyr His Gln Pro Pro Thr

Substitute Sequence Listing_USSN 10587032_PP019482.007

1	5	10	15																
Arg	Arg	Leu	Asn	Pro	Met	Ala	Pro	Pro	Ser	Arg	Gln	Gly	Ser	Glu	Leu				
			20					25					30						
His	Glu	Ile	Ile	Asp	Lys	Ser	Arg	Lys	Glu	Gly	Asp	Thr	Glu	Ala	Trp				
		35					40					45							
Gln	Phe	Pro	Val	Thr	Leu	Glu	Pro	Met	Pro	Pro	Gly	Glu	Gly	Ala	Gln				
	50					55					60								
Glu	Gly	Glu	Pro	Pro	Thr	Val	Glu	Ala	Arg	Tyr	Lys	Ser	Phe	Ser	Ile				
65					70					75					80				
Lys	Met	Leu	Lys	Asp	Met	Lys	Glu	Gly	Val	Lys	Gln	Tyr	Gly	Pro	Asn				
				85					90					95					
Ser	Pro	Tyr	Met	Arg	Thr	Leu	Leu	Asp	Ser	Ile	Ala	Tyr	Gly	His	Arg				
			100					105					110						
Leu	Ile	Pro	Tyr	Asp	Trp	Glu	Ile	Leu	Ala	Lys	Ser	Ser	Leu	Ser	Pro				
		115					120						125						
Ser	Gln	Phe	Leu	Gln	Phe	Lys	Thr	Trp	Trp	Ile	Asp	Gly	Val	Gln	Glu				
	130					135					140								
Gln	Val	Arg	Arg	Asn	Arg	Ala	Ala	Asn	Pro	Pro	Val	Asn	Ile	Asp	Ala				
145					150					155					160				
Asp	Gln	Leu	Leu	Gly	Ile	Gly	Gln	Asn	Trp	Ser	Thr	Ile	Ser	Gln	Gln				
				165					170					175					
Ala	Leu	Met	Gln	Asn	Glu	Ala	Ile	Glu	Gln	Val	Arg	Ala	Ile	Cys	Leu				
			180					185					190						
Arg	Ala	Trp	Glu	Lys	Ile	Gln	Asp	Pro	Gly	Ser	Thr	Cys	Pro	Ser	Phe				
		195					200					205							
Asn	Thr	Val	Arg	Gln	Gly	Ser	Lys	Glu	Pro	Tyr	Pro	Asp	Phe	Val	Ala				
	210					215					220								
Arg	Leu	Gln	Asp	Val	Ala	Gln	Lys	Ser	Ile	Ala	Asp	Glu	Lys	Ala	Gly				
225					230					235					240				
Lys	Val	Ile	Val	Glu	Leu	Met	Ala	Tyr	Glu	Asn	Ala	Asn	Pro	Glu	Cys				
				245					250					255					
Gln	Ser	Ala	Ile	Lys	Pro	Leu	Lys	Gly	Lys	Val	Pro	Ala	Gly	Ser	Asp				
			260					265					270						
Val	Ile	Ser	Glu	Tyr	Val	Lys	Ala	Cys	Asp	Gly	Ile	Gly	Gly	Ala	Met				
		275					280					285							
His	Lys	Ala	Met	Leu	Met	Ala	Gln	Ala	Ile	Thr	Gly	Val	Val	Leu	Gly				
	290					295					300								
Gly	Gln	Val	Arg	Thr	Phe	Gly	Gly	Lys	Cys	Tyr	Asn	Cys	Gly	Gln	Ile				
305					310					315					320				
Gly	His	Leu	Lys	Lys	Asn	Cys	Pro	Val	Leu	Asn	Lys	Gln	Asn	Ile	Thr				
				325					330					335					
Ile	Gln	Ala	Thr	Thr	Thr	Gly	Arg	Glu	Pro	Pro	Asp	Leu	Cys	Pro	Arg				

Cys Lys Lys Gly Lys His Trp Ala Ser Gln Cys Arg Ser Lys Phe Asp
355 360 365
Lys Asn Gly Gln Pro Leu Ser Gly Asn Glu Gln Arg Gly Gln Pro Gln
370 375 380
Ala Pro Gln Gln Thr Gly Ala Phe Pro Ile Gln Pro Phe Val Pro Gln
385 390 395 400
Gly Phe Gln Gly Gln Gln Pro Pro Leu Ser Gln Val Phe Gln Gly Ile
405 410 415
Ser Gln Leu Pro Gln Tyr Asn Asn Cys Pro Ser Pro Gln Ala Ala Val
420 425 430
Gln Gln

<210> 9
<211> 666
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 9
Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
1 5 10 15
Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
20 25 30
Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
35 40 45
Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly
50 55 60
Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65 70 75 80
Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
85 90 95
Thr Glu Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Ile
100 105 110
Ile Asp Cys Asn Glu Asn Thr Gly Lys Lys Ser Gln Lys Glu Thr Glu
115 120 125
Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
130 135 140
Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145 150 155 160
Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
165 170 175
Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
180 185 190
Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val

Substitute Sequence Listing_USSN 10587032_PP019482.007

195

200

205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
 210 215 220
 Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
 225 230 235 240
 Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
 245 250 255
 Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
 260 265 270
 Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
 275 280 285
 Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
 290 295 300
 Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
 305 310 315 320
 Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
 325 330 335
 Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
 340 345 350
 Gln Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
 355 360 365
 Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
 370 375 380
 Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln
 385 390 395 400
 Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
 405 410 415
 Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Glu Lys Ile Gln Asp
 420 425 430
 Pro Gly Ser Thr Cys Pro Ser Phe Asn Thr Val Arg Gln Gly Ser Lys
 435 440 445
 Glu Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Val Ala Gln Lys
 450 455 460
 Ser Ile Ala Asp Glu Lys Ala Arg Lys Val Ile Val Glu Leu Met Ala
 465 470 475 480
 Tyr Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys
 485 490 495
 Gly Lys Val Pro Ala Gly Ser Asp Val Ile Ser Glu Tyr Val Lys Ala
 500 505 510
 Cys Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln
 515 520 525
 Ala Ile Thr Gly Val Val Leu Gly Gly Gln Val Arg Thr Phe Gly Arg

Substitute Sequence Listing_USSN 10587032_PP019482.007

530

535

540

Lys Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Lys Asn Cys Pro
545 550 555 560

Val Leu Asn Lys Gln Asn Ile Thr Ile Gln Ala Thr Thr Thr Gly Arg
565 570 575

Glu Pro Pro Asp Leu Cys Pro Arg Cys Lys Lys Gly Lys His Trp Ala
580 585 590

Ser Gln Cys Arg Ser Lys Phe Asp Lys Asn Gly Gln Pro Leu Ser Gly
595 600 605

Asn Glu Gln Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
610 615 620

Pro Ile Gln Pro Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Pro
625 630 635 640

Leu Ser Gln Val Phe Gln Gly Ile Ser Gln Leu Pro Gln Tyr Asn Asn
645 650 655

Cys Pro Pro Pro Gln Ala Ala Val Gln Gln
660 665

<210> 10

<211> 1000

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 10

atgggcaacc	attgtcggga	aacgagcaaa	ggggccagcc	tcaggcccca	caacaaactg	60
gggcattccc	aattcagcca	tttgttcctc	agggttttca	gggacaacaa	ccccactgt	120
cccaagtgtt	tcagggaata	agccagttac	cacaatacaa	caattgtccc	ccgccacaag	180
cggcagtgca	gcagtagatt	tatgtactat	acaagcagtc	tctctgcttc	caggggagcc	240
cccacaaaaa	acccccacag	gggtatatgg	acccctgcct	aaggggactg	taggactaat	300
cttgggacga	tcaagtctaa	atctaaaagg	agttcaaatt	catactagt	tggttgattc	360
agactataaa	ggcgaaattc	aattggttat	tagctcttca	attccttgga	gtgccagtcc	420
aagagacagg	attgtcfaat	tattactcct	gccatacatt	aaggggtgga	atagtgaat	480
aaaaagaata	ggagggtctg	gaagcactga	tccaacagga	aaggctgcat	attgggcaag	540
tcagggtctca	gagaacagac	ctgtgtgtaa	ggccattatt	caaggaaaac	agtttgaagg	600
gttggtagac	actggagcag	atgtctctat	cattgcttta	aatcagtggc	caaaaaattg	660
gcctaaacaa	aaggctgtta	caggacttgt	cgcatagggc	acagcctcag	aagtgtatca	720
aagtacggag	attttacatt	gcttagggcc	agataatcaa	gaaagtactg	ttcagccaat	780
gattacttca	attcctctta	atctgtgggg	tcgagattta	ttacaacaat	ggggtgcgga	840
aatcaccatg	cccgtcccat	catatagccc	cacgagtcaa	aaaatcatga	ccaagatggg	900
atatatacca	ggaaagggac	tagggaaaaa	tgaagatggc	attaaaattc	cagttgaggc	960
taaaataaat	caagaaagag	aaggaatagg	gaatccttgc			1000

<210> 11

<211> 1004

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 11

atgggcaacc	attgtcggga	aacaggaaga	ggggccagcc	tcaggccccc	caacaaactg	60
gggcattccc	agttcaactg	tttgttcctc	agggttttca	aggacaacaa	cccctacaga	120
aaataccacc	acttcaggga	gtcagccaat	tacaacaatc	caacagctgt	cccgcgccac	180
agcaggcagc	gccacagtag	atttatgttc	cacccaaatg	gtctctttac	tccctggaga	240
gccccacaa	aagattccta	gaggggtata	tggcccgcgt	ccagaaggga	gggtaggcct	300
tattttaggg	agatcaagtc	taaatttgaa	gggagtccaa	attcatactg	gggtaattta	360
ttcagattat	aaagggggaa	ttcagttagt	gatcagctcc	actgttcctt	ggagtgccaa	420

tccaggtgat	agaattgctc	aattactgct	tttgcttat	gttaaaattg	gggaaaacaa	480
aacggaaaga	acaggagggt	ttggaagtac	caaccctgca	ggaaaagcca	cttattgggc	540
taatcaggtc	tcagaggata	gacccgtgtg	tacagtcact	attcagggaa	agagtttgaa	600
ggattagtgg	ataccaggc	tgatgtttct	atcatcggca	taggcaccgc	ctcagaagtg	660
tatcaaagtg	ccatgatttt	acattgtcta	ggatctgata	atcaagaaag	tacggttcag	720
cctatgatca	cttctatttc	aatcaattta	tggggccgag	acttgttaca	acaatggcat	780
gcagagatta	ctatcccgac	ctccctatac	agccccagga	atcaaaaaat	catgactaaa	840
atgggatagc	tccctaaaaa	gggactagga	agaaatgaag	atggcattaa	agtcccaact	900
gaggctgaaa	aaaatcaaaa	agagaaaagg	aatagggcat	cctttttaga	agcggtcact	960
gtagagcctc	caaaacccat	tccattaatt	tggggggaaa	aaaa		1004

<400>	12						
atggagattt	tacattgctt	agggccagat	aatcaagaaa	gtactgttca	gccaatgatt	60	
acttcaattc	ctcttaatct	gtgggggtcga	gattttattac	aacaatgggg	tgcggaaatc	120	
accattgcccg	ctccattata	tagccccacg	agtcaaaaaa	tcatgaccaa	gatgggatat	180	
ataaccaggaa	agggactagg	gaaaaaatgaa	gatggcatta	aagtccagt	tgaggctaaa	240	
ataaatcaag	aaagagaagg	aatagggtat	cctttttag			279	

<400> 13
Met Glu Ile Leu His Cys Leu Gly Pro Asp Asn Gln Glu Ser Thr Val
1 5 10 15
Gln Pro Met Ile Thr Ser Ile Pro Leu Asn Leu Trp Gly Arg Asp Leu
20 25 30
Leu Gln Gln Trp Gly Ala Glu Ile Thr Met Pro Ala Pro Leu Tyr Ser
35 40 45
Pro Thr Ser Gln Lys Ile Met Thr Lys Met Gly Tyr Ile Pro Gly Lys
50 55 60
Gly Leu Gly Lys Asn Glu Asp Gly Ile Lys Val Pro Val Glu Ala Lys
65 70 75 80
Ile Asn Gln Glu Arg Glu Gly Ile Gly Tyr Pro Phe
85 90

^{<400>} 14
 Trp Ala Thr Ile Val Gly Lys Arg Ala Lys Gly Pro Ala Ser Gly Pro
 1 5 10 15
 Thr Thr Asn Trp Gly Ile Pro Asn Ser Ala Ile Cys Ser Ser Gly Phe
 20 25 30
 Ser Gly Thr Thr Thr Pro Thr Val Pro Ser Val Ser Gly Asn Lys Pro
 35 40 45
 Val Thr Thr Ile Gln Gln Leu Ser Pro Ala Thr Ser Gly Ser Ala Ala

Substitute Sequence Listing_USSN 10587032_PP019482.007

50

55

60

Val Asp Leu Cys Thr Ile Gln Ala Val Ser Leu Leu Pro Gly Glu Pro
65 70 75 80

Pro Gln Lys Thr Pro Thr Gly Val Tyr Gly Pro Leu Pro Lys Gly Thr
85 90 95

Val Gly Leu Ile Leu Gly Arg Ser Ser Leu Asn Leu Lys Gly Val Gln
100 105 110

Ile His Thr Ser Val Val Asp Ser Asp Tyr Lys Gly Glu Ile Gln Leu
115 120 125

Val Ile Ser Ser Ser Ile Pro Trp Ser Ala Ser Pro Arg Asp Arg Ile
130 135 140

Ala Gln Leu Leu Leu Leu Pro Tyr Ile Lys Gly Gly Asn Ser Glu Ile
145 150 155 160

Lys Arg Ile Gly Gly Leu Gly Ser Thr Asp Pro Thr Gly Lys Ala Ala
165 170 175

Tyr Trp Ala Ser Gln Val Ser Glu Asn Arg Pro Val Cys Lys Ala Ile
180 185 190

Ile Gln Gly Lys Gln Phe Glu Gly Leu Val Asp Thr Gly Ala Asp Val
195 200 205

Ser Ile Ile Ala Leu Asn Gln Trp Pro Lys Asn Trp Pro Lys Gln Lys
210 215 220

Ala Val Thr Gly Leu Val Gly Ile Gly Thr Ala Ser Glu Val Tyr Gln
225 230 235 240

Ser Thr Glu Ile Leu His Cys Leu Gly Pro Asp Asn Gln Glu Ser Thr
245 250 255

Val Gln Pro Met Ile Thr Ser Ile Pro Leu Asn Leu Trp Gly Arg Asp
260 265 270

Leu Leu Gln Gln Trp Gly Ala Glu Ile Thr Met Pro Ala Pro Ser Tyr
275 280 285

Ser Pro Thr Ser Gln Lys Ile Met Thr Lys Met Gly Tyr Ile Pro Gly
290 295 300

Lys Gly Leu Gly Lys Asn Glu Asp Gly Ile Lys Ile Pro Val Glu Ala
305 310 315 320

Lys Ile Asn Gln Glu Arg Glu Gly Ile Gly Asn Pro Cys
325 330

<210> 15

<211> 2896

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 15

atggcattaa	aattccagtt	gaggctaaaa	taaatcaaga	aagagaagga	ataggggaatc	60
cttgctaggg	gcggccactg	tagagcctcc	taaaccata	ccattaactt	ggaaaacaga	120
aaaaccagtg	tgggtaaatc	agtggccgct	accaaaca	aaactggagg	ctttacattt	180
attagcaaat	gaacagttag	aaaagggtca	tattgagcct	tcgttctcac	cttgggaattc	240

Substitute Sequence Listing_USSN 10587032_PP019482.007

tcctgtgttt	gtaattcaga	agaaatcagg	caaattggcgt	atgttaactg	acttaagggc	300
tgtaaacgcc	gtaattcaac	ccatggggcc	tctccaaccc	gggttgccct	ctccggccat	360
gatcccaaaa	gattggcctt	taattataat	tgatctaaag	gattgctttt	ttaccatccc	420
tctggcagag	caggattgcg	aaaaatttgc	ctttactata	ccagccataa	ataataaaga	480
accagccacc	aggtttcagt	ggaaagtgtt	acctcagggg	atgcttaata	gtccaactat	540
ttgtcagact	ttttaggtgc	gagctcttca	accagttaga	gaaaagtttt	cagactgtta	600
tattattcat	tgtattgatg	atattttatg	tgctgcagaa	acgaaagata	aattaattga	660
ctgttataca	aattctgaag	cagaggttgc	caatgctgga	ctggcaatag	catctgataa	720
gatccaaacc	tctactcctt	ttcattattt	agggatgcag	atagaaaata	gaaaaattaa	780
gccacaaaaa	atagaaataa	gaaaagacac	attaaaaaca	ctaaatgatt	ttcaaaaatt	840
actaggagat	attaattgga	ttcggccaac	tctaggcatt	cctacttatg	ccatgtcaaa	900
tttgttctct	atcttaagag	gagactcaga	cttaaatagt	aaaagaatgt	taaccccaga	960
ggcaacaaaa	gaaattaaat	tagtgggaaga	aaaaattcag	tcagcgcaaa	taaatagaat	1020
agatccctta	gccccactcc	aactttttag	ttttgccact	gcacattctc	caacaggcat	1080
cattattcaa	aatactgacg	ttgtggagtg	gtcatttcctt	cctcacagta	cagttaagac	1140
ttttacattg	tacttgatc	aaatagctac	attaatcggg	cagacaagat	tacgaataat	1200
aaaattatgt	gggaatgacc	cagacaaaat	agttgtccct	ttaaccaagg	aacaagttag	1260
acaagccttt	atcaattctg	gtgcatggaa	gattggctct	gctaattttg	tggaattat	1320
tgataatcat	tacccaaaaa	caaagatctt	ccagttctta	aaattgacta	cttggaattct	1380
acctaataat	accagacgtg	aaccttttag	aaatgctcta	acagtattta	ctgatgggtc	1440
cagcaatgga	aaagcagctt	acacaggacc	gaaagaacga	gtaatcaaaa	ctccatatca	1500
atcggctcaa	agagcagagt	tggttgagct	cattacagtg	ttacaagatt	ttgaccaacc	1560
tatcaatatt	atatcagatt	ctgcatatgt	agtacaggct	acaagggatg	ttgagacagc	1620
tctaattaaa	tatagcatgg	atgatcagtt	aaaccagcta	ttcaatttat	tacaacaaac	1680
tgtaagaaaa	agaaatttcc	cattttatat	tacacatatt	cgagcacaca	ctaattttacc	1740
agggcctttg	actaaagcaa	atgaacaagc	tgacttactg	gtatcatctg	caactcataa	1800
agcacaagaa	cttcattgctt	tgactcatgt	aaatgcagca	ggattaaaaa	acaaatttga	1860
tgtcacatgg	aaacaggcaa	aagatattgt	acaacattgc	accagtgctc	aagtcttaca	1920
cctgcccact	caagaggcag	gagttaatcc	cagaggtctg	tgctctaatt	cattatggca	1980
aatggatgtc	acgcatgtac	cttcatttgg	aagattatca	tatgttcacg	taacagttga	2040
tacttattca	catttcatat	gggcaacttg	ccaaacagga	gaaagtactt	cccattgttaa	2100
aaaacattta	ttgtcttggt	ttgtctgtaat	gggagttcca	gaaaaaatca	aaactgacaa	2160
tggaaccagga	tattgtagta	aagctttcca	aaaattctta	agtcagtggg	aaatttcaca	2220
tacaacagga	attccttata	attcccaagg	acaggccata	gttgaaagaa	ctaatagaac	2280
actcaaaaact	caattagtta	aacaaaaaga	agggggagac	agtaaggagt	gtaccactcc	2340
tcagatgcaa	cttaatctag	cactctatac	tttaaatttt	ttaaacattt	atagaaatca	2400
gactactact	tctgcagaac	aacatcttac	tggtaaaaag	aacagcccac	atgaaggaaa	2460
actaatttgg	tggaagata	ataaaaaata	gacatgggaa	ataggggaagg	tgataacgtg	2520
ggggagaggt	tttgcttggt	tttcaccagg	agaaaatcag	cttccttggt	ggatacccac	2580
tgacatttgg	aagttctaca	atgaaccat	cagagatgca	aagaaaagca	cctccgcgga	2640
gacggagaca	tcgcaatcga	gcaccgttga	ctcacaagat	gaacaaaatg	gtgacgtcag	2700
agaacagat	gaagtggcca	tccaccaaga	aggcagagcc	gccaaacttg	gcacaactaa	2760
agaagctgac	gcagttagct	acaaaatatc	tagagaacac	aaaggtgaca	caaaccccag	2820
agagtatgct	gcttgagccc	ttgatgattg	tatcaatggt	ggtaagtctc	cctatgcctg	2880
caggagcagc	tgacg					2896

<210> 16

<211> 2619

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 16

atgttaactg	acttaagggc	tgtaaacgcc	gtaattcaac	ccatggggcc	tctccaaccc	60
gggttgccct	ctccggccat	gatcccaaaa	gattggcctt	taattataat	tgatctaaag	120
gattgctttt	ttaccatccc	tctggcagag	caggattgcg	aaaaatttgc	ctttactata	180
ccagccataa	ataataaaga	accagccacc	aggtttcagt	ggaaagtgtt	acctcagggg	240
atgcttaata	gtccaaactat	ttgtcagact	ttttaggtgc	gagctcttca	accagttaga	300
gaaaagtttt	cagactgtta	tattattcat	tgatttgatg	atattttatg	tgctgcagaa	360
acgaaagata	aattaattga	ctgttataca	tttctgcaag	cagaggttgc	caatgctgga	420
ctggcaatag	catctgataa	gatccaaacc	tctactcctt	ttcattattt	agggatgcag	480
atagaaaata	gaaaaattaa	gccacaaaaa	atagaaataa	gaaaagacac	attaaaaaca	540
ctaaatgatt	ttcaaaaatt	actaggagat	attaattgga	ttcggccaac	tctaggcatt	600
cctacttatg	ccatgtcaaa	tttgttctct	atcttaagag	gagactcaga	cttaaatagt	660

Substitute Sequence Listing_USSN 10587032_PP019482.007

aaaagaatgt	taaccccgaga	ggcaacaaaa	gaaattaaat	tagtggaga	aaaaattcag	720
tcagcgcaaa	taaatagaat	agatccccta	gccccactcc	aacttttgat	ttttgccact	780
gcacattctc	caacaggcat	cattattcaa	aatactgac	ttgtggagt	gtcattccct	840
cctcacagta	cagttaagac	ttttacattg	tacttggatc	aaatagctac	attaatcgg	900
cagacaagat	tacgaataat	aaaattatgt	gggaatgacc	cagacaaaat	agttgtccct	960
ttaaccaagg	aacaagttag	acaagccttt	atcaattctg	gtgcatggaa	gattgggtct	1020
gctaattttg	tgggaattat	tgataatcat	tacccaaaaa	caaagatctt	ccagttctta	1080
aaattgacta	cttggattct	acctaaaatt	accagacgtg	aacctttaga	aaatgctcta	1140
acagtattta	ctgatggttc	cagcaatgga	aaagcagctt	acacaggacc	gaaagaacga	1200
gtaatcaaaa	ctccatatca	atcggctcaa	agagcagagt	tggttgcagt	cattacagtg	1260
ttacaagatt	ttgaccaacc	tatcaatatt	atatcagatt	ctgcatatgt	agtacaggct	1320
acaagggatg	ttgagacagc	tctaattaaa	tatagcatgg	atgatcagtt	aaaccagcta	1380
ttcaatttat	tacaacaaac	tgtaagaaaa	agaaatttcc	cattttatat	tacacatatt	1440
cgagcacaca	ctaattttac	agggcctttg	actaaagcaa	atgaacaagc	tgacttactg	1500
gtatcatctg	cactcataaa	agcacaagaa	cttcattgctt	tgactcatgt	aaatgcagca	1560
ggattaaaaa	acaaatttga	tgtcacatgg	aaacaggcaa	aagatattgt	acaacattgc	1620
acccagtgtc	aagtccttaca	cctgcccact	caagaggcag	gagttaatcc	cagaggctctg	1680
tgctcctaag	cattatggca	aatggatgtc	acgcattgtac	cttcattttg	aagattatca	1740
tatgttcacg	taacagttga	tacttattca	catttcatat	gggcaacttg	ccaaacagga	1800
gaaagtactt	cccattgttaa	aaaacattta	ttgtcttggt	ttgctgtaat	gggagttcca	1860
gaaaaaatca	aaactgacaa	tggaaccagga	tattgtagta	aagctttcca	aaaattctta	1920
agtcagtgga	aaatttcaca	tacaacagga	attccttata	attcccaagg	acaggccata	1980
ggtgaaagaa	ctaatagaac	actcaaaact	caattagtta	aacaaaaaga	agggggagac	2040
agtaaggagt	gtaccactcc	tcagatgcaa	cttaattctag	cactctatac	tttaaatatt	2100
ttaaacattt	atagaaatca	gactactact	tctgcagaac	aacatcttac	tggtaaaaag	2160
aacagcccac	atgaaggaaa	actaatttgg	tggaagata	gtaaaaataa	gacatgggaa	2220
ataggggaag	tgataacgtg	ggggagaggt	tttgcttggt	tttcaccagg	agaaaatcag	2280
cttcctgttt	ggatacccac	tagacatttg	aagttctaca	atgaaccat	cagagatgca	2340
aagaaaagca	cctccgcgga	gacggagaca	tcgcaatcga	gcaccgttga	ctcacaagat	2400
gaacaaaatg	gtgacgtcag	agaacagat	gaagttgcca	tccaccaaga	aggcagagcc	2460
gccaacttgg	gcacaactaa	agaagctgac	gcagttagct	acaaaatatc	tagagaacac	2520
aaaggtgaca	caaacccag	agagtatgct	gcttgacgcc	ttgatgattg	tatcaatggt	2580
ggtaagtctc	cctatgcctg	caggagcagc	tgcagctaa			2619

<210> 17

<211> 2671

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 17

atggcattaa	agtcccaact	gaggctgaaa	aaaatcaaaa	aaagaaaagg	aatagggcat	60
ccitttttaga	agcggtcact	gtagagcctc	caaaacccat	tccattaatt	tggggggaaa	120
aaaaaaaactg	tatggtaaat	cagtagccgc	ttccaaaaca	aaaactggag	gctttacact	180
tattagcaaa	gaaacagtta	gaaaaaggac	atattgagcc	ttcattttcg	ccttggaatt	240
ctcctgtttg	taattcagaa	aaaatccggc	agatggcgta	tgctaactga	cttaagagcc	300
attaatgccca	taattcaacc	catggggggg	ctccccctcc	ggttgccctc	tccagccatg	360
gtcccccttta	attataattg	atctgaagga	ttgctttttt	accattcctc	tggtcaaaaga	420
ggatttttgaa	aaatttgctt	ttactatacc	agcctaaata	ataaagaacc	agccaccagg	480
tttcagtgga	aagtattgcc	tcaggggaatg	cttaataatt	caactatttg	tcagactttc	540
atagctcaag	ctctgcaacc	agttagagac	aagttttcag	actgttatat	cgttcattat	600
gttgatattt	tgtgtgctgc	agaaacgaga	gacaaattaa	ttgaccgtta	cacatttctc	660
agacagaggt	tgccaacgcg	ggactgacaa	tagcatctga	taagattcaa	acctctcctc	720
ctttccatta	cttggggaatg	caggtagagg	aaaggaaaat	taaaccacaa	aaaatagaaa	780
taagaaaaga	cacattaaaa	acattaaatg	agtttcaaaa	gttggttagga	gatactaatt	840
ggattcggag	atattaattg	gatttgGCCA	actctaggca	ttcctactta	tgccatgtca	900
attttgttct	ctttcttaag	aggggacttg	gaattaaata	gtgaaagaat	gttacctcca	960
gaggcaacta	aagaaattaa	attaattgaa	gaaaaaaatt	cggtcagcac	aagtaaatag	1020
gatcacttgg	ccccactcca	aattttgatt	tttggtactg	cacatttctc	aacagccatc	1080
attgttcaaa	acacagatct	tgtggatttg	tccttccttc	ctcatagtac	aattaagact	1140
tttacattgt	acttggatca	aatggctaca	ttaattggct	aggggaagatt	acgaataata	1200
acattgtgtg	gaaatgaccc	agataaaaatc	actgttcctt	tcaacaagca	acaagttaga	1260
caagccttta	tcagttcttg	tgcatggcag	attggtcttg	ctaattttct	gggaattatt	1320
gataatcatt	acccaaaaac	aaaaatcttc	cagttcttaa	aattgactac	ttggattcta	1380

Substitute Sequence Listing_USSN 10587032_PP019482.007

cctaaaatta	ccagacgtga	accttttagaa	aatgctctaa	cagtattttac	tgatggttcc	1440
agcaatggaa	aagcggctta	cacagggccg	aaagaacgag	taatcaaaac	tccgtatcaa	1500
tcagctcaaa	gagcagagtt	ggttgcagtc	attacagtg	tacaagattt	tgaccaacct	1560
atcaatatta	tatcagattc	tgcatatgta	gtacaggcta	caagggatgt	tgagacagct	1620
ctaattaaat	atagcacgga	cgatcattta	aaccagctat	tcaattttatt	acaacaaact	1680
gtaagaaaaa	gaaattttccc	attttatatt	actcatattc	gagcacacac	taattttacca	1740
gggcctttga	ctaaagcaaa	tgaacaagct	gacttactgg	tatcatctgc	attcataaaa	1800
gcacaagaac	ttcttgcttt	gactcatgta	aatgcagcag	gattaaaaaa	caaatttgat	1860
gtcacatgga	aacaggcaaa	agatattgta	caacattgca	cccagtgta	agtctttacac	1920
ctgtccactc	aagaggcagg	agttaatccc	agaggctctgt	gtcctaatgc	gttatggcaa	1980
atggatggca	cgcattgttc	ttcatttgga	agattatcat	atgttcatgt	aacagttgat	2040
acttattcac	atttcatatg	ggcaacttgc	caaacaggag	aaagtacttc	ccatgtttaa	2100
aaacattttat	tatcttggtt	tgctgtaattg	ggagttccag	aaaaaatcaa	aactgacaat	2160
ggaccaggat	attgtagtta	agctttccaa	aaattcttaa	gtcagtggaa	aatttcacat	2220
acaacaggaa	ttccttataa	ttcccaagga	caggccatag	ttgaaagaac	taatagaaca	2280
ctcaaaactc	aattagttaa	acaaaaagaa	gggggagaca	gtaaggagt	taccactcct	2340
cagatgcaac	ttaatctagc	actctatact	ttaaattttt	taaacattta	tagaaatcag	2400
actactactt	ctgcaaaaaca	acatcttact	ggtaaaaagc	acagcccaca	tgaaggaaaa	2460
ctaatttggt	ggaaagataa	taaaaaataag	acatgggaaa	tagggaaggt	gataacgtgg	2520
gggagagggt	ttgcttgtag	ttcaccagga	gaaaatcagg	ttcctgtttg	gataccct	2580
agacatttga	agttctacaa	tgaaccctac	ggagatgcaa	agaaaagggc	ctccacagag	2640
atggttaacc	cagtcacatg	gatggataat	c			2671

<210> 18
 <211> 4086
 <212> DNA
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 18						
atggggcctc	tccaaccg	gttgccctct	ccggccatga	tcccaaaaga	ttggccttta	60
attataattg	atctaaagga	ttgctttttt	accatccctc	tggcagagca	ggattgtgaa	120
aaatttgcct	ttactatacc	agccataaat	ataaagaac	cagccaccag	gtttcagtgg	180
aaagtgttac	ctcagggaat	gcttaatagt	ccaactat	gtcagacttt	tgtaggtcga	240
gctcttcaac	cagtgaagaa	aaagttttca	gactgttata	ttattcatta	tattgatgat	300
attttatgtg	ctgcagaaac	gaaagataaa	ttaattgact	gttatacatt	tctgcaagca	360
gaggttgcca	atgctggact	ggcaatagca	tccgataaga	tccaaacctc	tactcctttt	420
cattattttg	ggatgcagat	agaaaataga	aaaatttaagc	cacaaaaaat	agaaataaga	480
aaagacacat	taaaaacact	aaatgatttt	caaaaattac	taggagatat	taattggatt	540
cggccaactc	taggcattcc	tacttatgcc	atgtcaaat	tgttctctat	cttaagagga	600
gactcagact	taaatagtca	agaatatta	acccagagg	caacaaaaga	aattaaatta	660
gtggaagaaa	aaattcagtc	agcgcaata	aatagaatag	atcccttagc	cccactccaa	720
cttttgattt	ttgccactgc	acattctcca	acaggcatca	ttattcaaaa	tactgatctt	780
gtggagtggg	cattccttcc	tcacagtaca	gttaagactt	ttacattgta	cttggatcaa	840
atagctacat	taatcggtca	gacaagatta	cgaataacaa	aattatgtgg	aaatgaccca	900
gacaaaatag	ttgtcccttt	aaccaaggaa	caagttagac	aagcctttat	caattctggg	960
gcatggcaga	ttggtcttgc	taattttgtg	ggacttattg	ataatcatta	cccaaaaaca	1020
aagatcttcc	agttcttaaa	attgactact	tggattctac	ctaaaattac	cagacgtgaa	1080
cctttagaaa	atgctctaac	agtatttact	gatggttcca	gcaatggaaa	agcagcttac	1140
acagggccga	aagaacgagt	aatcaaaact	ccatatcaat	cggctcaaag	agacgagttg	1200
gttgacgtca	ttacagtgtt	acaagatttt	gaccaacctc	tcaatattat	atcagattct	1260
gcatatgtag	tacaggctac	aagggatgtt	gagacagctc	taattaaata	tagcatggat	1320
gatcagttaa	accagctatt	caatttatta	caacaaactg	taagaaaaag	aaatttccca	1380
ttttatatta	cttatattcg	agcacacact	aattttaccag	ggcctttgac	taaagcaa	1440
gaacaagctg	acttactggg	atcatctgca	ctcataaaag	cacaagaact	tcatgctttg	1500
actcatgtaa	atgcagcagg	attaaaaaac	aaattttgat	tcacatggaa	acaggcaaaa	1560
gatattgtac	aacattgcac	ccagtgtaaa	gtcttacacc	tgcccactca	agaggcagga	1620
gttaatccca	gaggtctgtg	tcctaattgc	ttatggcaaa	tggatgtcac	gcatgtacct	1680
tcatttgga	gattatcata	tgttcatgta	acagttgata	cttattcaca	tttcatatgg	1740
gcaacttgcc	aaacaggaga	aagtacttcc	catgttaaaa	aacattttatt	gtcttggttt	1800
gctgtaatgg	gagttccaga	aaaaatcaaa	actgacaatg	gaccaggata	ttgtagtaaa	1860
gctttccaaa	aattcttaag	tcagtggaaa	atttcacata	caacaggaat	tccttataat	1920
tcccaaggac	aggccatagt	tgaagaagac	aatagaacac	tcaaaactca	attagttaaa	1980
caaaaagaag	ggggagacag	taaggagtgt	accactcttc	agatgcaact	taatctagca	2040

Substitute Sequence Listing_USSN 10587032_PP019482.007

ctctatactt	taaatttttt	aaacattttat	agaaatcaga	ctactacttc	tgcagaacaa	2100
catcttactg	gtaaaaagaa	cagcccacat	gaaggaaaac	taatttggtg	gaaagataat	2160
aaaaataaga	catgggaaat	aggggaagtg	ataacgtggg	ggagagggtt	tgcttggtt	2220
tcaccaggag	aaaatcagct	tcctgtttgg	ttaccacta	gacatttgaa	gttctacaat	2280
gaacccatcg	gagatgcaaa	gaaaagggcc	tccacggaga	tggtaacacc	agtcacatgg	2340
atggataatc	ctatagaagt	atatgttaat	gatagtatat	gggtacctgg	ccccatagat	2400
gatcgctgcc	ctgccaacc	tgaggaagaa	gggatgatga	taaatatttc	cattgggtat	2460
cgttatcctc	ctatttgcct	agggagagca	ccaggatggt	taatgcctgc	agtcctcaaat	2520
tggttggtag	aagtacctac	tgtcagtccc	atcagtagat	tcacttatca	catggtaagc	2580
gggatgtcac	tcaggccacg	ggtaaattat	ttacaagact	tttcttatca	aagatcatta	2640
aaatttagac	ctaaagggaa	accttgcccc	aaggaaattc	caaagaatc	aaaaaatata	2700
gaagttttag	tttggaaga	atgtgtggcc	aatagtgcgg	tgatattata	aaacaatgaa	2760
tttggaacta	ttatagattg	ggcacctcga	ggtcaattct	accacaattg	ctcaggacaa	2820
actcagtcgt	gtccaagtgc	acaagtgagt	ccagctgttg	atagcgactt	aacagaaagt	2880
ttagacaaac	ataagcataa	aaaattgcag	tctttctacc	cttgggaatg	gggagaaaaa	2940
ggaatctcta	ccccaagacc	aaaaatagta	agtcctgttt	ctggctctga	acatccagaa	3000
ttatggaggc	ttactgtggc	ctcacaccac	attagaattt	ggtctggaaa	tcaaacttta	3060
gaaacaagag	attgtaagcc	attttatact	gtcgacctaa	attccagctc	aacagttcct	3120
ttacaaagtt	gcgtaaagcc	cccttatatg	ctagttgtag	gaaatatagt	tattaaacca	3180
gactcccgag	ctataacctg	tgaaaattgt	agattgctta	cttgcatgga	ttcaactttt	3240
aatgtggcaac	accgtattct	gctggtgaga	gcaagagagg	gcgtgtggat	ccctgtgtcc	3300
atggaccgac	cgtgggaggc	ctcaccatcc	gtccatattt	tgactgaagt	attaaaagg	3360
gttttaata	gatccaaaag	attcattttt	actttaattg	cagtgattat	gggattaatt	3420
gcagtcacag	ctacggctgc	tgtagcagga	gttgcattgc	actcttctgt	tcagtcagta	3480
aactttgtta	atgattggca	aaagaattct	acaagattgt	ggaattcaca	atctagtatt	3540
gatcaaaaa	tggaacaatc	aattaatgat	cttagacaaa	ctgtcatttg	gatgggagac	3600
agactcatga	gcttagaaca	tcgtttccag	ttacaatgtg	actggaatac	gtcagatttt	3660
tgtattacac	cccaaattta	taatgagtct	gagcatcact	gggacatgg	tagacgcat	3720
ctacagggaa	gagaagataa	tctcacttta	gacatttcca	aattaaaaga	acaaattttc	3780
gaagcatcaa	aagcccattt	aaatttggtg	ccaggaactg	aggcaattgc	aggagttgct	3840
gatggcctcg	caaattctta	ccctgtcact	tggtgtaaga	ccattggaag	tacatcgatt	3900
ataaatctca	tattaatcct	tgtgtgcctg	ttttgtctgt	tgtagtctg	caggtgtacc	3960
caacagctcc	gaagagacag	cgaccatcga	gaacgggcca	tgatgacgat	ggcgggtttg	4020
tcgaaaagaa	aagggggaaa	tgtggggaaa	agcaagagag	atcaaattgt	tactgtgtct	4080
gtgtag						4086

<210> 19

<211> 872

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 19

Met	Leu	Thr	Asp	Leu	Arg	Ala	Val	Asn	Ala	Val	Ile	Gln	Pro	Met	Gly
1				5					10					15	

Pro	Leu	Gln	Pro	Gly	Leu	Pro	Ser	Pro	Ala	Met	Ile	Pro	Lys	Asp	Trp
		20						25					30		

Pro	Leu	Ile	Ile	Ile	Asp	Leu	Lys	Asp	Cys	Phe	Phe	Thr	Ile	Pro	Leu
		35				40						45			

Ala	Glu	Gln	Asp	Cys	Glu	Lys	Phe	Ala	Phe	Thr	Ile	Pro	Ala	Ile	Asn
	50					55					60				

Asn	Lys	Glu	Pro	Ala	Thr	Arg	Phe	Gln	Trp	Lys	Val	Leu	Pro	Gln	Gly
65					70				75					80	

Met	Leu	Asn	Ser	Pro	Thr	Ile	Cys	Gln	Thr	Phe	Val	Gly	Arg	Ala	Leu
				85					90					95	

Gln	Pro	Val	Arg	Glu	Lys	Phe	Ser	Asp	Cys	Tyr	Ile	Ile	His	Cys	Ile
			100					105					110		

Substitute Sequence Listing_USSN 10587032_PP019482.007

Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp Lys Leu Ile Asp Cys
 115 120 125
 Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala Gly Leu Ala Ile Ala
 130 135 140
 Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His Tyr Leu Gly Met Gln
 145 150 155 160
 Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile Glu Ile Arg Lys Asp
 165 170 175
 Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu Leu Gly Asp Ile Asn
 180 185 190
 Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr Ala Met Ser Asn Leu
 195 200 205
 Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn Ser Lys Arg Met Leu
 210 215 220
 Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val Glu Glu Lys Ile Gln
 225 230 235 240
 Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala Pro Leu Gln Leu Leu
 245 250 255
 Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile Ile Ile Gln Asn Thr
 260 265 270
 Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser Thr Val Lys Thr Phe
 275 280 285
 Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile Gly Gln Thr Arg Leu
 290 295 300
 Arg Ile Ile Lys Leu Cys Gly Asn Asp Pro Asp Lys Ile Val Val Pro
 305 310 315 320
 Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile Asn Ser Gly Ala Trp
 325 330 335
 Lys Ile Gly Leu Ala Asn Phe Val Gly Ile Ile Asp Asn His Tyr Pro
 340 345 350
 Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr Thr Trp Ile Leu Pro
 355 360 365
 Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala Leu Thr Val Phe Thr
 370 375 380
 Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly Pro Lys Glu Arg
 385 390 395 400
 Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Ala Glu Leu Val Ala
 405 410 415
 Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile Asn Ile Ile Ser
 420 425 430
 Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val Glu Thr Ala Leu
 435 440 445

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu Phe Asn Leu Leu
450 455 460

Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr Ile Thr His Ile
465 470 475 480

Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys Ala Asn Glu Gln
485 490 495

Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala Gln Glu Leu His
500 505 510

Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn Lys Phe Asp Val
515 520 525

Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys Thr Gln Cys Gln
530 535 540

Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn Pro Arg Gly Leu
545 550 555 560

Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His Val Pro Ser Phe
565 570 575

Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr Tyr Ser His Phe
580 585 590

Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser His Val Lys Lys
595 600 605

His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro Glu Lys Ile Lys
610 615 620

Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe Gln Lys Phe Leu
625 630 635 640

Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro Tyr Asn Ser Gln
645 650 655

Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu Lys Thr Gln Leu
660 665 670

Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu Cys Thr Thr Pro Gln
675 680 685

Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe Leu Asn Ile Tyr
690 695 700

Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His Leu Thr Gly Lys Lys
705 710 715 720

Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys Asp Ser Lys Asn
725 730 735

Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp Gly Arg Gly Phe Ala
740 745 750

Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp Ile Pro Thr Arg
755 760 765

His Leu Lys Phe Tyr Asn Glu Pro Ile Arg Asp Ala Lys Lys Ser Thr
770 775 780

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ser Ala Glu Thr Glu Thr Ser Gln Ser Ser Thr Val Asp Ser Gln Asp
785 790 795 800
Glu Gln Asn Gly Asp Val Arg Arg Thr Asp Glu Val Ala Ile His Gln
805 810 815
Glu Gly Arg Ala Ala Asn Leu Gly Thr Thr Lys Glu Ala Asp Ala Val
820 825 830
Ser Tyr Lys Ile Ser Arg Glu His Lys Gly Asp Thr Asn Pro Arg Glu
835 840 845
Tyr Ala Ala Cys Ser Leu Asp Asp Cys Ile Asn Gly Gly Lys Ser Pro
850 855 860
Tyr Ala Cys Arg Ser Ser Cys Ser
865 870

<210> 20
<211> 1361
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<220>
<221> SITE
<222> 917
<223> Xaa is any amino acid

<400> 20
Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met Ile Pro Lys
1 5 10 15
Asp Trp Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe Phe Thr Ile
20 25 30
Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr Ile Pro Ala
35 40 45
Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys Val Leu Pro
50 55 60
Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe Val Gly Arg
65 70 75 80
Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr Ile Ile His
85 90 95
Tyr Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp Lys Leu Ile
100 105 110
Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala Gly Leu Ala
115 120 125
Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His Tyr Leu Gly
130 135 140
Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile Glu Ile Arg
145 150 155 160
Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu Leu Gly Asp
165 170 175
Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr Ala Met Ser

Substitute Sequence Listing_USSN 10587032_PP019482.007

<p>Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn Ser Gln Arg 195</p> <p>Ile Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val Glu Glu Lys 210</p> <p>Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala Pro Leu Gln 225</p> <p>Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile Ile Ile Gln 245</p> <p>Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser Thr Val Lys 260</p> <p>Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile Gly Gln Thr 275</p> <p>Arg Leu Arg Ile Thr Lys Leu Cys Gly Asn Asp Pro Asp Lys Ile Val 290</p> <p>Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile Asn Ser Gly 305</p> <p>Ala Trp Gln Ile Gly Leu Ala Asn Phe Val Gly Leu Ile Asp Asn His 325</p> <p>Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr Thr Trp Ile 340</p> <p>Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala Leu Thr Val 355</p> <p>Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly Pro Lys 370</p> <p>Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Asp Glu Leu 385</p> <p>Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile Asn Ile 405</p> <p>Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val Glu Thr 420</p> <p>Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu Phe Asn 435</p> <p>Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr Ile Thr 450</p> <p>Tyr Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys Ala Asn 465</p> <p>Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala Gln Glu 485</p> <p>Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn Lys Phe 500</p> <p>Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys Thr Gln</p>	<p>180</p> <p>185</p> <p>200</p> <p>215</p> <p>230</p> <p>245</p> <p>265</p> <p>280</p> <p>295</p> <p>310</p> <p>330</p> <p>345</p> <p>360</p> <p>375</p> <p>390</p> <p>405</p> <p>425</p> <p>440</p> <p>455</p> <p>470</p> <p>490</p> <p>505</p>	<p>190</p> <p>205</p> <p>220</p> <p>235</p> <p>250</p> <p>270</p> <p>285</p> <p>300</p> <p>315</p> <p>335</p> <p>350</p> <p>365</p> <p>380</p> <p>395</p> <p>410</p> <p>430</p> <p>445</p> <p>460</p> <p>475</p> <p>495</p> <p>510</p>
--	---	--

Substitute Sequence Listing_USSN 10587032_PP019482.007

515		520		525	
Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn Pro Arg					
530		535		540	
Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His Val Pro					
545		550		555	560
Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr Tyr Ser					
	565		570		575
His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser His Val					
	580		585		590
Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro Glu Lys					
	595	600		605	
Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe Gln Lys					
	610	615		620	
Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro Tyr Asn					
625	630		635		640
Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu Lys Thr					
	645	650		655	
Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu Cys Thr Thr					
	660	665		670	
Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe Leu Asn					
	675	680		685	
Ile Tyr Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His Leu Thr Gly					
	690	695		700	
Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys Asp Asn					
705	710		715		720
Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp Gly Arg Gly					
	725	730		735	
Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp Leu Pro					
	740	745		750	
Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Gly Asp Ala Lys Lys					
	755	760		765	
Arg Ala Ser Thr Glu Met Val Thr Pro Val Thr Trp Met Asp Asn Pro					
	770	775		780	
Ile Glu Val Tyr Val Asn Asp Ser Ile Trp Val Pro Gly Pro Ile Asp					
785	790		795		800
Asp Arg Cys Pro Ala Lys Pro Glu Glu Glu Gly Met Met Ile Asn Ile					
	805	810		815	
Ser Ile Gly Tyr Arg Tyr Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly					
	820	825		830	
Cys Leu Met Pro Ala Val Gln Asn Trp Leu Val Glu Val Pro Thr Val					
	835	840		845	
Ser Pro Ile Ser Arg Phe Thr Tyr His Met Val Ser Gly Met Ser Leu					

Substitute Sequence Listing_USSN 10587032_PP019482.007

850

855

860

Arg Pro Arg Val Asn Tyr Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu
865 870 875 880

Lys Phe Arg Pro Lys Gly Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu
885 890 895

Ser Lys Asn Thr Glu Val Leu Val Trp Glu Glu Cys Val Ala Asn Ser
900 905 910

Ala Val Ile Leu Xaa Asn Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala
915 920 925

Pro Arg Gly Gln Phe Tyr His Asn Cys Ser Gly Gln Thr Gln Ser Cys
930 935 940

Pro Ser Ala Gln Val Ser Pro Ala Val Asp Ser Asp Leu Thr Glu Ser
945 950 955 960

Leu Asp Lys His Lys His Lys Lys Leu Gln Ser Phe Tyr Pro Trp Glu
965 970 975

Trp Gly Glu Lys Gly Ile Ser Thr Pro Arg Pro Lys Ile Val Ser Pro
980 985 990

Val Ser Gly Pro Glu His Pro Glu Leu Trp Arg Leu Thr Val Ala Ser
995 1000 1005

His His Ile Arg Ile Trp Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp
1010 1015 1020

Cys Lys Pro Phe Tyr Thr Val Asp Leu Asn Ser Ser Leu Thr Val Pro
1025 1030 1035 1040

Leu Gln Ser Cys Val Lys Pro Pro Tyr Met Leu Val Val Gly Asn Ile
1045 1050 1055

Val Ile Lys Pro Asp Ser Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu
1060 1065 1070

Leu Thr Cys Ile Asp Ser Thr Phe Asn Trp Gln His Arg Ile Leu Leu
1075 1080 1085

Val Arg Ala Arg Glu Gly Val Trp Ile Pro Val Ser Met Asp Arg Pro
1090 1095 1100

Trp Glu Ala Ser Pro Ser Val His Ile Leu Thr Glu Val Leu Lys Gly
1105 1110 1115 1120

Val Leu Asn Arg Ser Lys Arg Phe Ile Phe Thr Leu Ile Ala Val Ile
1125 1130 1135

Met Gly Leu Ile Ala Val Thr Ala Thr Ala Ala Val Ala Gly Val Ala
1140 1145 1150

Leu His Ser Ser Val Gln Ser Val Asn Phe Val Asn Asp Trp Gln Lys
1155 1160 1165

Asn Ser Thr Arg Leu Trp Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu
1170 1175 1180

Ala Asn Gln Ile Asn Asp Leu Arg Gln Thr Val Ile Trp Met Gly Asp

Substitute Sequence Listing_USSN 10587032_PP019482.007

1185 1190 1195 1200
 Arg Leu Met Ser Leu Glu His Arg Phe Gln Leu Gln Cys Asp Trp Asn
 1205 1210 1215
 Thr Ser Asp Phe Cys Ile Thr Pro Gln Ile Tyr Asn Glu Ser Glu His
 1220 1225 1230
 His Trp Asp Met Val Arg Arg His Leu Gln Gly Arg Glu Asp Asn Leu
 1235 1240 1245
 Thr Leu Asp Ile Ser Lys Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys
 1250 1255 1260
 Ala His Leu Asn Leu Val Pro Gly Thr Glu Ala Ile Ala Gly Val Ala
 1265 1270 1275 1280
 Asp Gly Leu Ala Asn Leu Asn Pro Val Thr Trp Val Lys Thr Ile Gly
 1285 1290 1295
 Ser Thr Ser Ile Ile Asn Leu Ile Leu Ile Leu Val Cys Leu Phe Cys
 1300 1305 1310
 Leu Leu Leu Val Cys Arg Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp
 1315 1320 1325
 His Arg Glu Arg Ala Met Met Thr Met Ala Val Leu Ser Lys Arg Lys
 1330 1335 1340
 Gly Gly Asn Val Gly Lys Ser Lys Arg Asp Gln Ile Val Thr Val Ser
 1345 1350 1355 1360
 Val

<210> 21
 <211> 956
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 21
 Asn Lys Ser Arg Lys Arg Arg Asn Arg Glu Ser Leu Leu Gly Ala Ala
 1 5 10 15
 Thr Val Glu Pro Pro Lys Pro Ile Pro Leu Thr Trp Lys Thr Glu Lys
 20 25 30
 Pro Val Trp Val Asn Gln Trp Pro Leu Pro Lys Gln Lys Leu Glu Ala
 35 40 45
 Leu His Leu Leu Ala Asn Glu Gln Leu Glu Lys Gly His Ile Glu Pro
 50 55 60
 Ser Phe Ser Pro Trp Asn Ser Pro Val Phe Val Ile Gln Lys Lys Ser
 65 70 75 80
 Gly Lys Trp Arg Met Leu Thr Asp Leu Arg Ala Val Asn Ala Val Ile
 85 90 95
 Gln Pro Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met Ile
 100 105 110
 Pro Lys Asp Trp Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe Phe
 Page 25

125

Page 26

Substitute Sequence Listing_USSN 10587032_PP019482.007

450

455

460

Thr Val Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly
 465 470 475 480
 Pro Lys Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Ala
 485 490 495
 Glu Leu Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile
 500 505 510
 Asn Ile Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val
 515 520 525
 Glu Thr Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu
 530 535 540
 Phe Asn Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr
 545 550 555 560
 Ile Thr His Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys
 565 570 575
 Ala Asn Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala
 580 585 590
 Gln Glu Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn
 595 600 605
 Lys Phe Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys
 610 615 620
 Thr Gln Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn
 625 630 635 640
 Pro Arg Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His
 645 650 655
 Val Pro Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr
 660 665 670
 Tyr Ser His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser
 675 680 685
 His Val Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro
 690 695 700
 Glu Lys Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe
 705 710 715 720
 Gln Lys Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro
 725 730 735
 Tyr Asn Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu
 740 745 750
 Lys Thr Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu Cys
 755 760 765
 Thr Thr Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe
 770 775 780
 Leu Asn Ile Tyr Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His Leu

785				790				795				800			
Thr	Gly	Lys	Lys	Asn 805	Ser	Pro	His	Glu	Gly 810	Lys	Leu	Ile	Trp	Trp 815	Lys
Asp	Asn	Lys	Asn 820	Lys	Thr	Trp	Glu	Ile 825	Gly	Lys	Val	Ile	Thr 830	Trp	Gly
Arg	Gly	Phe 835	Ala	Cys	Val	Ser	Pro 840	Gly	Glu	Asn	Gln	Leu 845	Pro	Val	Trp
Ile	Pro 850	Thr	Arg	His	Leu	Lys 855	Phe	Tyr	Asn	Glu	Pro 860	Ile	Arg	Asp	Ala
Lys 865	Lys	Ser	Thr	Ser	Ala 870	Glu	Thr	Glu	Thr	Ser 875	Gln	Ser	Ser	Thr	Val 880
Asp	Ser	Gln	Asp	Glu 885	Gln	Asn	Gly	Asp	Val 890	Arg	Arg	Thr	Asp	Glu 895	Val
Ala	Ile	His	Gln 900	Glu	Gly	Arg	Ala	Ala 905	Asn	Leu	Gly	Thr	Thr 910	Lys	Glu
Ala	Asp	Ala 915	Val	Ser	Tyr	Lys	Ile 920	Ser	Arg	Glu	His	Lys 925	Gly	Asp	Thr
Asn	Pro 930	Arg	Glu	Tyr	Ala	Ala 935	Cys	Ser	Leu	Asp	Asp 940	Cys	Ile	Asn	Gly
Gly 945	Lys	Ser	Pro	Tyr	Ala 950	Cys	Arg	Ser	Ser	Cys 955	Ser				

```
<210> 22
<211> 2000
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)
```

<400>	22					
atgaacccat	cagagatgca	aagaaaagca	cctccgcgga	gacggagaca	tcgcaatcga	60
gcaccgttga	ctcacaagat	gaacaaaatg	gtgacgtcag	agaacagat	gaagttgcca	120
tccaccaaga	aggcagagcc	gccaaacttg	gcacaactaa	agaagctgac	gcagtttagct	180
acaaaatatc	tagagaacac	aaaggtgaca	caaaacccag	agagtatgct	gcttgcacgc	240
ttgatgattg	tatcaatggg	ggtaagtctc	cctatgcctg	caggagcagc	tgcagtaaac	300
tatacctact	gggcctatgt	gcctttcccg	cccttaattc	gggcagtcac	atggatggat	360
aatcctacag	aagtatatgt	taatgatagt	gtatgggtac	ctggcccccat	agatgatcgc	420
tgccctgcca	aacctgagga	agaagggatg	atgataaata	tttccattgg	gtatcattat	480
cctcctatit	gcctagggag	agcaccagga	tgtttaatgc	ctgcagtcga	aaattgggtg	540
gtagaagtac	ctactgtcag	tcccatctgt	agattcactt	atcacatggt	aagcgggatg	600
tcactcaggc	cacggcgtaaa	ttattttcaa	gacttttctt	atcaaatagc	attaaaattt	660
agacctaaag	ggaaaccttg	ccccaaggaa	attcccaaaag	aatcaaaaaa	tacagaagtt	720
ttagtttggg	aagaatgtgt	ggccaatagt	gcggtgatat	tacaaaacaa	tgaattcgga	780
actattatag	attgggcacc	tcgaggtcaa	ttctaccaca	attgctcagg	acaaactcag	840
tcgtgtccaa	gtgcacaagt	gagtccagct	gttgatagcg	acttaacaga	aagtttagac	900
aaacataaag	ataaaaaaatt	gcagtctttc	tacccttggg	aatggggaga	aaaagggaatc	960
tctatcccaa	gacccaaaaat	agtaagtcct	gtttcttggtc	ctgaacatcc	agaattatgg	1020
aggcttactg	tggccttcaca	ccacattaga	atttggtctg	gaaatcaaac	tttagaaaca	1080
agagatcgta	agccatttta	tactattgac	ctgaattcca	gtctaacagt	tcctttacaa	1140
agttgcgtaa	agccccctta	tatgctagtt	gtaggaaata	tagttattaa	accagactcc	1200
cagactataa	cctgtgaaaa	ttgtagattg	cttacttgca	ttgattcaac	ttttaattgg	1260
caacaccgta	ttctgctggt	gagagcaaga	gagggcgtgt	ggatccctgt	gtccatggac	1320
cgaccgtggg	aggcctcgcc	atccgtccat	attttgactg	aagtattaaa	aggtgtttta	1380
aaatagatcca	aaagattcat	ttttacttta	attgcagctg	ttagtgggatt	aattgctattc	1440
acagctacgg	ctgctgtagc	aggagttgca	ttgcactctt	ctgttcaqtc	agtaaaccttt	1500

Substitute Sequence Listing_USSN 10587032_PP019482.007

gttaatgatt	ggcaaaaaaa	ttctacaaga	ttgtggaatt	cacaatctag	tattgatcaa	1560
aaattggcaa	atcaaattaa	tgatcttaga	caaactgtca	tttggatggg	agacagactc	1620
atgagcttag	aacatcgttt	ccagttacaa	tgtgactgga	atacgtcaga	tttttgtatt	1680
acaccccaaa	tttataatga	gtctgagcat	cactgggaca	tggttagacg	ccatctacag	1740
ggaagagaag	ataatctcac	tttagacatt	tccaaattaa	aagaacaaat	tttcgaagca	1800
tcaaaagccc	atttaaattt	ggtgccagga	actgaggcaa	ttgcaggagt	tgctgatggc	1860
ctcgcaaatc	ttaaccctgt	cacttgggtt	aagaccattg	gaagtactac	gattataaat	1920
ctcatattaa	tccctgtgtg	cctgttttgt	ctgttgtag	tctgcagggt	tacccaacag	1980
ctccgaagag	acagcgacca					2000

<210> 23

<211> 2085

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 23

atgcaaagaa	aagcacctcc	gcggagacgg	agacatcgca	atcgagcacc	gttgactcac	60
aagatgaaca	aaatggtgac	gtcagaagaa	cagatgaagt	tgccatccac	caagaaggca	120
gagccgcaa	cttgggcaca	actaaagaag	ctgacgcagt	tagctacaaa	atatctagag	180
aacacaaagg	tgacacaaac	cccagagagt	atgctgcttg	cagccttgat	gattgtatca	240
atggtggtaa	gtctccctat	gcctgcagga	gcagctgcag	ctaactatac	ctactgggcc	300
tatgtgcctt	tcccgccttt	aattcgggca	gtcacatgga	tgataaatcc	tacagaagta	360
tatgttaatg	atagtgtatg	ggtacctggc	cccatagatg	atcgctgccc	tgccaaacct	420
gaggaagaag	ggatgatgat	aaatatttcc	attgggtatc	attatcctcc	tatttgcccta	480
gggagagcac	caggatgttt	aatgcctgca	gtccaaaatt	ggttggtaga	agtacctact	540
gtcagtccca	tctgtagatt	cacttatcac	atggtgaagc	ggatgtcact	caggccacgg	600
gtaaatttatt	tacaagactt	ttcttatcaa	agatcattaa	aatttagacc	taaagggaaa	660
ccttgcccca	aggaaattcc	caaagaatca	aaaaatacag	aagttttagt	ttgggaagaa	720
tgtgtggcca	atagtgcggt	gatattacaa	aacaatgaat	tcggaactat	tatagattgg	780
gcacctcgag	gtcaattcta	ccacaattgc	tcaggacaaa	ctcagtcgtg	tcaaagtgc	840
caagtgagtc	cagctgttga	tagcgactta	acagaaagtt	tagacaaaca	taagcataaa	900
aaattgcagt	ctttctaccc	ttgggaatgg	ggagaaaaag	gaatctctac	cccaagacca	960
aaaatagtaa	gtctctgttc	tggtcctgaa	catccagaat	tatggaggct	tactgtggcc	1020
tcacaccaca	ttagaatttg	gtctggaaat	caaactttag	aaacaagaga	tcgtaagcca	1080
ttttatacta	ttgacctgaa	ttccagtcta	acagttcctt	tacaaagttg	cgtaaagccc	1140
ccttatatgc	tagttgtagg	aaatatagtt	attaaccag	actcccagac	tataacctgt	1200
gaaaattgta	gattgcttac	ttgcattgat	tcaactttta	attggcaaca	ccgtattctg	1260
ctggtgagag	caagagaggg	cgtgtggatc	cctgtgtcca	tggaaccgacc	gtgggaggcc	1320
tcgccatccg	tccatatttt	gactgaagta	ttaaaagggt	ttttaaatag	atccaaaaga	1380
ttcattttta	ctttaattgc	agtgattatg	ggattaattg	cagtcacagc	tacggctgct	1440
gtagcaggag	ttgcattgca	ctcttctgtt	cagtcagtaa	actttgttaa	tgattggcaa	1500
aaaaattcta	caagattgtg	gaattcacaa	tctagtattg	atcaaaaatt	ggcaaatcaa	1560
attaatgata	ttagacaaac	tgatcatttg	atgggagaca	gactcatgag	cttagaacat	1620
cgtttccagt	tacaattgtg	ctggaatacg	tcagattttt	gtattacacc	ccaaatttat	1680
aatgactctg	agcatcactg	ggacatggtt	agacgccatc	tacagggaag	agaagataat	1740
ctcactttag	acatttccaa	attaaaagaa	caaattttcg	aagcatcaaa	agcccattta	1800
aatttgggtg	caggaactga	ggcaattgca	ggagtgtctg	atggcctcgc	aaatcttaac	1860
cctgtcactt	gggttaagac	cattggaagt	actacgatta	taaatctcat	attaatcctt	1920
gtgtgcctgt	tttgtctgtt	gttagtctgc	aggtgtaccc	aacagctccg	aagagacagc	1980
gaccatcgag	aacgggccat	gatgacgatg	gcggttttgt	cgaaaagaaa	agggggaaat	2040
gtggggaaaa	gcaagagaga	tcagattgtt	actgtgtctg	tgtag		2085

<210> 24

<211> 1665

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 24

gtcacatgga	tggataatcc	tatagaagta	tatgttaatg	atagtgtatg	ggtacctggc	60
cccacagatg	atcgctgccc	tgccaaacct	gaggaagaag	ggatgatgat	aaatatttcc	120
attgtgtatc	gttatcctcc	tatttgccta	gggagagcac	caggatgttt	aatgcctgca	180
gtccaaaatt	ggttggtaga	agtacctact	gtcagtccta	acagtagatt	cacttatcac	240
atggtgaagc	ggatgtcact	caggccacgg	gtaaaattatt	tacaagactt	ttcttatcaa	300

Substitute Sequence Listing_USSN 10587032_PP019482.007

agatcattaa	aatttagacc	taaagggaaa	ccttgcccca	aggaaattcc	caaagaatca	360
aaaaatacag	aagttttagt	ttgggaagaa	tgtgtggcca	atagtgcggt	gatattacaa	420
aacaatgaat	tcggaactat	tatagattgg	gcacctcgag	gtcaattcta	ccacaattgc	480
tcaggacaaa	ctcagtcgtg	tccaagtgca	caagtgaagc	cagctgttga	tagcgactta	540
acagaaagtc	tagacaaaca	taagcataaa	aaattacagt	ctttctaccc	ttgggaatgg	600
ggagaaaaag	gaatctctac	cccaagacca	gaaataataa	gtcctgtttc	tggctctgaa	660
catccagaat	tatggaggct	ttggcctgac	accacattag	aatttggtct	ggaaatcaaa	720
ctttagaAAC	aagagatcgt	aagccatttt	atactatcga	cctaaattcc	agtctaacgg	780
ttcctttaca	aagttgcgta	aagccctctt	atatgctagt	tgtaggaaat	atagttatta	840
aaccagactc	ccaaactata	acctgtgaaa	attgtagatt	gtttacttgc	attgattcaa	900
cttttaattg	gcggcaccgt	attctgctgg	tgagagcaag	agagggcggt	tggatctctg	960
tgtccgtgga	ctgaccgtgg	gaggcctcgc	catccatcca	tattttgact	gaagtattaa	1020
aagacatttt	aaatagatcc	aaaagattca	tttttacctt	aattgcagtg	attatgggat	1080
taattgcagt	cacagctacg	gctgctgtgg	caggagttag	attgcactct	tctgttcagt	1140
cggtaaactt	tgtaaatgat	tggaacaaag	attctacaag	attgtggaat	tcacaatcta	1200
gtattgatca	aaaattggca	aatcaaatca	atgatcttag	acaaactgtc	atttggatgg	1260
gagacagact	catgagctta	gaacattggt	tccagttaca	gtgtgactgg	aatacgtcag	1320
atTTTTgtat	tacaccccaa	atttataatg	agtctgagca	tcactgggac	atgggttagac	1380
gccatctaca	gggaagagaa	gataatctca	ctttagacat	ttccaaatta	aaataacaaa	1440
ttttcgaagc	atcaaaagcc	catttaaagt	ttagtccagc	aactgaggca	attgcaggag	1500
ttgctgatgg	cctcgcaaat	cttaaccctg	tcacttgggt	taagaccatc	ggaagtacta	1560
tgattataaa	tctcatatta	atccttgtgt	gcctgttttg	tctgttggtt	gtctgcaggt	1620
gtaccaaca	gctccgaaga	gacagcgacc	atcgagaacg	ggcca		1665

<210> 25

<211> 4086

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 25

atggggcctc	tccaacccgg	gttgccctct	cgggccatga	tcccaaaaga	ttggccttta	60
attataattg	atctaaagga	ttgctttttt	accatccctc	tggcagagca	ggattgtgaa	120
aaatttgcct	ttactatacc	agccataaat	aataaagaac	cagccaccag	gtttcagttg	180
aaagtgttac	ctcaggggaat	gcttaatagt	ccaactatct	gtcagacttt	tgtaggtcga	240
gctcttcaac	cagtgaagaa	aaagtgttca	gactgtttata	ttattcatta	tattgatgat	300
atTTTTatgtg	ctgcagaaac	gaaagataaa	ttatttgact	gttatacatt	tctgcaagca	360
gaggttgcca	atgctggact	ggcaatagca	tccgataaga	tccaaacctc	tactcctttt	420
cattattttg	ggatgcagat	agaaaataga	aaaattaagc	cacaaaaaat	agaaaataga	480
aaagacacat	taaaaaacct	aaatgatttt	caaaaattac	taggagatat	taattggatt	540
cggccaactc	taggcattcc	tacttatgcc	atgtcaaatt	tgttctctat	cttaagagga	600
gactcagact	taaatagtca	aagaatatta	acccagagg	caacaaaaga	aattaaatta	660
gtggaagaaa	aaattcagtc	agcgcaaat	aatagaatag	atcccttagc	cccactccaa	720
cttttgattt	ttgccactgc	acattctcca	acaggcatca	ttattcaaaa	tactgatctt	780
gtggagtggg	cattccttcc	tcacagtaca	gttaagactt	ttacattgta	cttggatcaa	840
atagctacat	taaatcggtc	gacaagatta	cgaataacaa	aattatgtgg	aaatgaccca	900
gacaaaatag	ttgtcccttt	aaccaaggaa	caagttagac	aagcctttat	caattctggt	960
gcatggcaga	ttggctcttg	taattttgtg	ggacttattg	ataatcatta	cccaaaaaca	1020
aagatcttcc	agttcttaaa	attgactact	tggattctac	ctaaaattac	cagacgtgaa	1080
ccittagaaa	atgctctaac	agtatttact	gatggttcca	gcaatggaaa	agcagcttac	1140
acagggccga	aagaacgagt	aatcaaaact	ccatatcaat	cggctcaaag	agacgagttg	1200
gttgacgtca	ttacagtgtt	acaagatttt	gaccaacct	tcaatattat	atcagattct	1260
gcatatgtag	tacaggctac	aagggtgtgt	gagacagctc	taattaaata	tagcatggat	1320
gatcagttaa	accagctatt	caatttatta	caacaaactg	taagaaaaag	aaatttccca	1380
ttttatatta	cttatattcg	agcacacact	aatttaccag	ggcctttgac	taaagcaaat	1440
gaacaagctg	acttactggg	atcatctgca	ctcataaaag	cacaagaact	tcatgctttg	1500
actcatgtaa	atgcagcagg	attaaaaaac	aaatttgatg	tcacatggaa	acaggcaaaa	1560
gatattgtac	aacattgcac	ccagtgtcaa	gtcttacacc	tgcccactca	agaggcagga	1620
gttaatccca	gaggtctgtg	tcctaattgca	ttatggcaaa	tggatgtcac	gcatgtacct	1680
tcatttgtaa	gattatcata	tgttcatgta	acagttgata	cttattcaca	tttcatatgg	1740
gcaacttgcc	aaacaggaga	aagtacttcc	catgttaaaa	aacattttatt	gtcttgtttt	1800
gctgtaatgg	gagttccaga	aaaaatcaaa	actgacaatg	gaccaggata	ttgtagtaaa	1860
gctttccaaa	aattcttaag	tcagtggaaa	atttcacata	caacagggaat	tccttataat	1920
tcccaaggac	aggccatagt	tgaagaact	aatagaacac	tcaaaactca	attagttaaa	1980

Substitute Sequence Listing_USSN 10587032_PP019482.007

caaaaagaag	ggggagacag	taaggagtgt	accactcctc	agatgcaact	taatctagca	2040
ctctatactt	taaatttttt	aaacattttat	agaaatcaga	ctactacttc	tgcagaacaa	2100
catcttactg	gtaaaaagaa	cagcccat	gaaggaaaac	taatttggtg	gaaagataat	2160
aaaaataaga	catgggaaat	aggggaaggtg	ataacgtggg	ggagagggtt	tgcttgtgtt	2220
tcaccaggag	aaaatcagct	tcctgttttg	ttaccacta	gacatttgaa	gttctacaat	2280
gaacccatcg	gagatgcaaa	gaaaagggcc	tccacggaga	tggtaacacc	agtcacatgg	2340
atggataatc	ctatagaagt	atatgttaat	gatagtatat	gggtacctgg	ccccatagat	2400
gatcgctgcc	ctgccaaacc	tgaggaagaa	gggatgatga	taaatatttc	cattgggtat	2460
cgttatcctc	ctatttgcct	aggagagca	ccaggatggt	taatgcctgc	agtccaaat	2520
tggttggtag	aagtacctac	tgtcagtccc	atcagtagat	tcacttatca	catggtaagc	2580
gggatgtcac	tcaggccacg	ggtaaattat	ttacaagact	tttcttatca	aagatcatta	2640
aaatttagac	ctaaagggaa	accttgcccc	aaggaaattc	ccaaagaatc	aaaaaataca	2700
gaagttttag	tttggaaga	atgtgtggcc	aatagtgcgg	tgatattata	aaacaatgaa	2760
tttggaacta	ttatagattg	ggcacctcga	ggccaattct	accacaattg	ctcaggacaa	2820
actcagtcgt	gtccaagtgc	acaagtga	ccagctgttg	atagcgactt	aacagaaagt	2880
ttagacaaac	ataagcataa	aaaattgcag	tctttctacc	cttggaatg	gggagaaaaa	2940
ggaatctcta	ccccaaagacc	aaaaatagta	agtcctgttt	ctggctcctga	acatccagaa	3000
ttatggaggc	ttactgtggc	ctcacaccac	attagaattt	ggctctggaaa	tcaaacttta	3060
gaaacaagag	attgtaagcc	atttttatact	gtcgacctaa	attccagtct	aacagttcct	3120
ttactaaagt	cggtaaagcc	cccttatatg	ctagttgtag	gaaatatagt	tattaaacca	3180
gactcccaga	ctataacctg	tgaaaattgt	agattgtcta	cttgcatgga	ttcaactttt	3240
aattggcaac	accgtattct	gctgggtgaga	gcaagagagg	gcgtgtggat	ccctgtgtcc	3300
atggaccgac	cgtgggaggc	ctcaccatcc	gtccatattt	tgactgaagt	attaaaagg	3360
gttttaata	gatccaaaag	attcattttt	actttaattg	cagtgattat	gggattaatt	3420
gcagtcacag	ctacggctgc	tgtagcagga	gttgcatg	actcttctgt	tcagtcagta	3480
aactttgtta	atgattggca	aaagaattct	acaagattgt	ggaattcaca	atctagtatt	3540
gatcaaaaat	tggcaaatca	aattaatgat	cttagacaaa	ctgtcatttg	gatgggagac	3600
agactcatga	gcttagaaca	tcgtttccag	ttacaatgtg	actggaatac	gtcagatttt	3660
tgtattacac	cccaaattta	taatgagtct	gagcatcact	gggacatgg	tagacgccat	3720
ctacagggaa	gagaagataa	tctcacttta	gacatttcca	aattaaaaga	acaaattttc	3780
gaagcatcaa	aagcccattt	aaatttggtg	ccaggaactg	aggcaattgc	aggagttgct	3840
gatggcctcg	caaatcttaa	ccctgtcact	tgggttaaga	ccattggaag	tacatcgatt	3900
ataaatctca	tattaatcct	tgtgtgcctg	ttttgtctgt	tgtagtctg	caggtgtacc	3960
caacagctcc	gaagagacag	cgaccatcga	gaacgggcca	tgatgacgat	ggcggttttg	4020
tcgaaaagaa	aagggggaaa	tgtggggaaa	agcaagagag	atcaaattgt	tactgtgtct	4080
gtgtag						4086

<210> 26
 <211> 694
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 26
 Met Gln Arg Lys Ala Pro Pro Arg Arg Arg His Arg Asn Arg Ala
 1 5 10 15
 Pro Leu Thr His Lys Met Asn Lys Met Val Thr Ser Glu Glu Gln Met
 20 25 30
 Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro Thr Trp Ala Gln Leu
 35 40 45
 Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu Glu Asn Thr Lys Val
 50 55 60
 Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala Leu Met Ile Val Ser
 65 70 75 80
 Met Val Val Ser Leu Pro Met Pro Ala Gly Ala Ala Ala Asn Tyr
 85 90 95
 Thr Tyr Trp Ala Tyr Val Pro Phe Pro Pro Leu Ile Arg Ala Val Thr
 100 105 110

Substitute Sequence Listing_USSN 10587032_PP019482.007

Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn Asp Ser Val Trp Val
 115 120 125
 Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys Pro Glu Glu Glu Gly
 130 135 140
 Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr Pro Pro Ile Cys Leu
 145 150 155 160
 Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val Gln Asn Trp Leu Val
 165 170 175
 Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe Thr Tyr His Met Val
 180 185 190
 Ser Gly Met Ser Leu Arg Pro Arg Val Asn Tyr Leu Gln Asp Phe Ser
 195 200 205
 Tyr Gln Arg Ser Leu Lys Phe Arg Pro Lys Gly Lys Pro Cys Pro Lys
 210 215 220
 Glu Ile Pro Lys Glu Ser Lys Asn Thr Glu Val Leu Val Trp Glu Glu
 225 230 235 240
 Cys Val Ala Asn Ser Ala Val Ile Leu Gln Asn Asn Glu Phe Gly Thr
 245 250 255
 Ile Ile Asp Trp Ala Pro Arg Gly Gln Phe Tyr His Asn Cys Ser Gly
 260 265 270
 Gln Thr Gln Ser Cys Gln Ser Ala Gln Val Ser Pro Ala Val Asp Ser
 275 280 285
 Asp Leu Thr Glu Ser Leu Asp Lys His Lys His Lys Lys Leu Gln Ser
 290 295 300
 Phe Tyr Pro Trp Glu Trp Gly Glu Lys Gly Ile Ser Thr Pro Arg Pro
 305 310 315 320
 Lys Ile Val Ser Pro Val Ser Gly Pro Glu His Pro Glu Leu Trp Arg
 325 330 335
 Leu Thr Val Ala Ser His His Ile Arg Ile Trp Ser Gly Asn Gln Thr
 340 345 350
 Leu Glu Thr Arg Asp Arg Lys Pro Phe Tyr Thr Ile Asp Leu Asn Ser
 355 360 365
 Ser Leu Thr Val Pro Leu Gln Ser Cys Val Lys Pro Pro Tyr Met Leu
 370 375 380
 Val Val Gly Asn Ile Val Ile Lys Pro Asp Ser Gln Thr Ile Thr Cys
 385 390 395 400
 Glu Asn Cys Arg Leu Leu Thr Cys Ile Asp Ser Thr Phe Asn Trp Gln
 405 410 415
 His Arg Ile Leu Leu Val Arg Ala Arg Glu Gly Val Trp Ile Pro Val
 420 425 430
 Ser Met Asp Arg Pro Trp Glu Ala Ser Pro Ser Val His Ile Leu Thr
 435 440 445

Substitute Sequence Listing_USSN 10587032_PP019482.007

Glu Val Leu Lys Gly Val Leu Asn Arg Ser Lys Arg Phe Ile Phe Thr
450 455 460
Leu Ile Ala Val Ile Met Gly Leu Ile Ala Val Thr Ala Thr Ala Ala
465 470 475 480
Val Ala Gly Val Ala Leu His Ser Ser Val Gln Ser Val Asn Phe Val
485 490 495
Asn Asp Trp Gln Lys Asn Ser Thr Arg Leu Trp Asn Ser Gln Ser Ser
500 505 510
Ile Asp Gln Lys Leu Ala Asn Gln Ile Asn Asp Leu Arg Gln Thr Val
515 520 525
Ile Trp Met Gly Asp Arg Leu Met Ser Leu Glu His Arg Phe Gln Leu
530 535 540
Gln Cys Asp Trp Asn Thr Ser Asp Phe Cys Ile Thr Pro Gln Ile Tyr
545 550 555 560
Asn Glu Ser Glu His His Trp Asp Met Val Arg Arg His Leu Gln Gly
565 570 575
Arg Glu Asp Asn Leu Thr Leu Asp Ile Ser Lys Leu Lys Glu Gln Ile
580 585 590
Phe Glu Ala Ser Lys Ala His Leu Asn Leu Val Pro Gly Thr Glu Ala
595 600 605
Ile Ala Gly Val Ala Asp Gly Leu Ala Asn Leu Asn Pro Val Thr Trp
610 615 620
Val Lys Thr Ile Gly Ser Thr Thr Ile Ile Asn Leu Ile Leu Ile Leu
625 630 635 640
Val Cys Leu Phe Cys Leu Leu Leu Val Cys Arg Cys Thr Gln Gln Leu
645 650 655
Arg Arg Asp Ser Asp His Arg Glu Arg Ala Met Met Thr Met Ala Val
660 665 670
Leu Ser Lys Arg Lys Gly Gly Asn Val Gly Lys Ser Lys Arg Asp Gln
675 680 685
Ile Val Thr Val Ser Val
690

<210> 27
<211> 1361
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<220>
<221> SITE
<222> 917
<223> Xaa is any amino acid

<400> 27
Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met Ile Pro Lys
1 5 10 15

Substitute Sequence Listing_USSN 10587032_PP019482.007

Asp	Trp	Pro	Leu	Ile	Ile	Ile	Asp	Leu	Lys	Asp	Cys	Phe	Phe	Thr	Ile		
			20					25					30				
Pro	Leu	Ala	Glu	Gln	Asp	Cys	Glu	Lys	Phe	Ala	Phe	Thr	Ile	Pro	Ala		
		35					40					45					
Ile	Asn	Asn	Lys	Glu	Pro	Ala	Thr	Arg	Phe	Gln	Trp	Lys	Val	Leu	Pro		
	50					55					60						
Gln	Gly	Met	Leu	Asn	Ser	Pro	Thr	Ile	Cys	Gln	Thr	Phe	Val	Gly	Arg		
65					70					75				80			
Ala	Leu	Gln	Pro	Val	Arg	Glu	Lys	Phe	Ser	Asp	Cys	Tyr	Ile	Ile	His		
				85					90					95			
Tyr	Ile	Asp	Asp	Ile	Leu	Cys	Ala	Ala	Glu	Thr	Lys	Asp	Lys	Leu	Ile		
			100					105					110				
Asp	Cys	Tyr	Thr	Phe	Leu	Gln	Ala	Glu	Val	Ala	Asn	Ala	Gly	Leu	Ala		
		115					120					125					
Ile	Ala	Ser	Asp	Lys	Ile	Gln	Thr	Ser	Thr	Pro	Phe	His	Tyr	Leu	Gly		
	130					135					140						
Met	Gln	Ile	Glu	Asn	Arg	Lys	Ile	Lys	Pro	Gln	Lys	Ile	Glu	Ile	Arg		
145					150					155					160		
Lys	Asp	Thr	Leu	Lys	Thr	Leu	Asn	Asp	Phe	Gln	Lys	Leu	Leu	Gly	Asp		
				165					170					175			
Ile	Asn	Trp	Ile	Arg	Pro	Thr	Leu	Gly	Ile	Pro	Thr	Tyr	Ala	Met	Ser		
			180					185					190				
Asn	Leu	Phe	Ser	Ile	Leu	Arg	Gly	Asp	Ser	Asp	Leu	Asn	Ser	Gln	Arg		
		195					200					205					
Ile	Leu	Thr	Pro	Glu	Ala	Thr	Lys	Glu	Ile	Lys	Leu	Val	Glu	Glu	Lys		
	210					215					220						
Ile	Gln	Ser	Ala	Gln	Ile	Asn	Arg	Ile	Asp	Pro	Leu	Ala	Pro	Leu	Gln		
225					230					235					240		
Leu	Leu	Ile	Phe	Ala	Thr	Ala	His	Ser	Pro	Thr	Gly	Ile	Ile	Ile	Gln		
				245					250				255				
Asn	Thr	Asp	Leu	Val	Glu	Trp	Ser	Phe	Leu	Pro	His	Ser	Thr	Val	Lys		
			260					265					270				
Thr	Phe	Thr	Leu	Tyr	Leu	Asp	Gln	Ile	Ala	Thr	Leu	Ile	Gly	Gln	Thr		
		275					280					285					
Arg	Leu	Arg	Ile	Thr	Lys	Leu	Cys	Gly	Asn	Asp	Pro	Asp	Lys	Ile	Val		
	290					295					300						
Val	Pro	Leu	Thr	Lys	Glu	Gln	Val	Arg	Gln	Ala	Phe	Ile	Asn	Ser	Gly		
305					310					315					320		
Ala	Trp	Gln	Ile	Gly	Leu	Ala	Asn	Phe	Val	Gly	Leu	Ile	Asp	Asn	His		
				325					330					335			
Tyr	Pro	Lys	Thr	Lys	Ile	Phe	Gln	Phe	Leu	Lys	Leu	Thr	Thr	Trp	Ile		
			340					345					350				

Substitute Sequence Listing_USSN 10587032_PP019482.007

Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala Leu Thr Val
 355 360 365
 Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly Pro Lys
 370 375 380
 Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Asp Glu Leu
 385 390 395 400
 Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile Asn Ile
 405 410 415
 Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val Glu Thr
 420 425 430
 Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu Phe Asn
 435 440 445
 Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr Ile Thr
 450 455 460
 Tyr Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys Ala Asn
 465 470 475 480
 Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala Gln Glu
 485 490 495
 Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn Lys Phe
 500 505 510
 Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys Thr Gln
 515 520 525
 Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn Pro Arg
 530 535 540
 Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His Val Pro
 545 550 555 560
 Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr Tyr Ser
 565 570 575
 His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser His Val
 580 585 590
 Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro Glu Lys
 595 600 605
 Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe Gln Lys
 610 615 620
 Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro Tyr Asn
 625 630 635 640
 Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu Lys Thr
 645 650 655
 Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu Cys Thr Thr
 660 665 670
 Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe Leu Asn
 675 680 685

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ile Tyr Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His Leu Thr Gly
690 695 700

Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys Asp Asn
705 710 715 720

Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp Gly Arg Gly
725 730 735

Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp Leu Pro
740 745 750

Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Gly Asp Ala Lys Lys
755 760 765

Arg Ala Ser Thr Glu Met Val Thr Pro Val Thr Trp Met Asp Asn Pro
770 775 780

Ile Glu Val Tyr Val Asn Asp Ser Ile Trp Val Pro Gly Pro Ile Asp
785 790 795 800

Asp Arg Cys Pro Ala Lys Pro Glu Glu Glu Gly Met Met Ile Asn Ile
805 810 815

Ser Ile Gly Tyr Arg Tyr Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly
820 825 830

Cys Leu Met Pro Ala Val Gln Asn Trp Leu Val Glu Val Pro Thr Val
835 840 845

Ser Pro Ile Ser Arg Phe Thr Tyr His Met Val Ser Gly Met Ser Leu
850 855 860

Arg Pro Arg Val Asn Tyr Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu
865 870 875 880

Lys Phe Arg Pro Lys Gly Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu
885 890 895

Ser Lys Asn Thr Glu Val Leu Val Trp Glu Glu Cys Val Ala Asn Ser
900 905 910

Ala Val Ile Leu Xaa Asn Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala
915 920 925

Pro Arg Gly Gln Phe Tyr His Asn Cys Ser Gly Gln Thr Gln Ser Cys
930 935 940

Pro Ser Ala Gln Val Ser Pro Ala Val Asp Ser Asp Leu Thr Glu Ser
945 950 955 960

Leu Asp Lys His Lys His Lys Lys Leu Gln Ser Phe Tyr Pro Trp Glu
965 970 975

Trp Gly Glu Lys Gly Ile Ser Thr Pro Arg Pro Lys Ile Val Ser Pro
980 985 990

Val Ser Gly Pro Glu His Pro Glu Leu Trp Arg Leu Thr Val Ala Ser
995 1000 1005

His His Ile Arg Ile Trp Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp
1010 1015 1020

Substitute Sequence Listing_USSN 10587032_PP019482.007

Cys Lys Pro Phe Tyr Thr Val Asp Leu Asn Ser Ser Leu Thr Val Pro
 1025 1030 1035 1040
 Leu Gln Ser Cys Val Lys Pro Pro Tyr Met Leu Val Val Gly Asn Ile
 1045 1050 1055
 Val Ile Lys Pro Asp Ser Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu
 1060 1065 1070
 Leu Thr Cys Ile Asp Ser Thr Phe Asn Trp Gln His Arg Ile Leu Leu
 1075 1080 1085
 Val Arg Ala Arg Glu Gly Val Trp Ile Pro Val Ser Met Asp Arg Pro
 1090 1095 1100
 Trp Glu Ala Ser Pro Ser Val His Ile Leu Thr Glu Val Leu Lys Gly
 1105 1110 1115 1120
 Val Leu Asn Arg Ser Lys Arg Phe Ile Phe Thr Leu Ile Ala Val Ile
 1125 1130 1135
 Met Gly Leu Ile Ala Val Thr Ala Thr Ala Ala Val Ala Gly Val Ala
 1140 1145 1150
 Leu His Ser Ser Val Gln Ser Val Asn Phe Val Asn Asp Trp Gln Lys
 1155 1160 1165
 Asn Ser Thr Arg Leu Trp Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu
 1170 1175 1180
 Ala Asn Gln Ile Asn Asp Leu Arg Gln Thr Val Ile Trp Met Gly Asp
 1185 1190 1195 1200
 Arg Leu Met Ser Leu Glu His Arg Phe Gln Leu Gln Cys Asp Trp Asn
 1205 1210 1215
 Thr Ser Asp Phe Cys Ile Thr Pro Gln Ile Tyr Asn Glu Ser Glu His
 1220 1225 1230
 His Trp Asp Met Val Arg Arg His Leu Gln Gly Arg Glu Asp Asn Leu
 1235 1240 1245
 Thr Leu Asp Ile Ser Lys Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys
 1250 1255 1260
 Ala His Leu Asn Leu Val Pro Gly Thr Glu Ala Ile Ala Gly Val Ala
 1265 1270 1275 1280
 Asp Gly Leu Ala Asn Leu Asn Pro Val Thr Trp Val Lys Thr Ile Gly
 1285 1290 1295
 Ser Thr Ser Ile Ile Asn Leu Ile Leu Ile Leu Val Cys Leu Phe Cys
 1300 1305 1310
 Leu Leu Leu Val Cys Arg Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp
 1315 1320 1325
 His Arg Glu Arg Ala Met Met Thr Met Ala Val Leu Ser Lys Arg Lys
 1330 1335 1340
 Gly Gly Asn Val Gly Lys Ser Lys Arg Asp Gln Ile Val Thr Val Ser
 1345 1350 1355 1360

Val

<210> 28
 <211> 699
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)
 <400> 28
 Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
 1 5 10 15
 His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
 20 25 30
 Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
 35 40 45
 Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
 50 55 60
 Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
 65 70 75 80
 Leu Met Ile Val Ser Met Val Val Ser Leu Pro Met Pro Ala Gly Ala
 85 90 95
 Ala Ala Ala Asn Tyr Thr Tyr Trp Ala Tyr Val Pro Phe Pro Pro Leu
 100 105 110
 Ile Arg Ala Val Thr Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn
 115 120 125
 Asp Ser Val Trp Val Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys
 130 135 140
 Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr
 145 150 155 160
 Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val
 165 170 175
 Gln Asn Trp Leu Val Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe
 180 185 190
 Thr Tyr His Met Val Ser Gly Met Ser Leu Arg Pro Arg Val Asn Tyr
 195 200 205
 Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu Lys Phe Arg Pro Lys Gly
 210 215 220
 Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu Ser Lys Asn Thr Glu Val
 225 230 235 240
 Leu Val Trp Glu Glu Cys Val Ala Asn Ser Ala Val Ile Leu Gln Asn
 245 250 255
 Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala Pro Arg Gly Gln Phe Tyr
 260 265 270
 His Asn Cys Ser Gly Gln Thr Gln Ser Cys Pro Ser Ala Gln Val Ser
 275 280 285

Substitute Sequence Listing_USSN 10587032_PP019482.007

Pro Ala Val Asp Ser Asp Leu Thr Glu Ser Leu Asp Lys His Lys His
290 295 300

Lys Lys Leu Gln Ser Phe Tyr Pro Trp Glu Trp Gly Glu Lys Gly Ile
305 310 315 320

Ser Thr Pro Arg Pro Lys Ile Val Ser Pro Val Ser Gly Pro Glu His
325 330 335

Pro Glu Leu Trp Arg Leu Thr Val Ala Ser His His Ile Arg Ile Trp
340 345 350

Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp Arg Lys Pro Phe Tyr Thr
355 360 365

Ile Asp Leu Asn Ser Ser Leu Thr Val Pro Leu Gln Ser Cys Val Lys
370 375 380

Pro Pro Tyr Met Leu Val Val Gly Asn Ile Val Ile Lys Pro Asp Ser
385 390 395 400

Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu Leu Thr Cys Ile Asp Ser
405 410 415

Thr Phe Asn Trp Gln His Arg Ile Leu Leu Val Arg Ala Arg Glu Gly
420 425 430

Val Trp Ile Pro Val Ser Met Asp Arg Pro Trp Glu Ala Ser Pro Ser
435 440 445

Val His Ile Leu Thr Glu Val Leu Lys Gly Val Leu Asn Arg Ser Lys
450 455 460

Arg Phe Ile Phe Thr Leu Ile Ala Val Ile Met Gly Leu Ile Ala Val
465 470 475 480

Thr Ala Thr Ala Ala Val Ala Gly Val Ala Leu His Ser Ser Val Gln
485 490 495

Ser Val Asn Phe Val Asn Asp Trp Gln Lys Asn Ser Thr Arg Leu Trp
500 505 510

Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu Ala Asn Gln Ile Asn Asp
515 520 525

Leu Arg Gln Thr Val Ile Trp Met Gly Asp Arg Leu Met Ser Leu Glu
530 535 540

His Arg Phe Gln Leu Gln Cys Asp Trp Asn Thr Ser Asp Phe Cys Ile
545 550 555 560

Thr Pro Gln Ile Tyr Asn Glu Ser Glu His His Trp Asp Met Val Arg
565 570 575

Arg His Leu Gln Gly Arg Glu Asp Asn Leu Thr Leu Asp Ile Ser Lys
580 585 590

Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys Ala His Leu Asn Leu Val
595 600 605

Pro Gly Thr Glu Ala Ile Ala Gly Val Ala Asp Gly Leu Ala Asn Leu
610 615 620

Substitute Sequence Listing_USSN 10587032_PP019482.007

Asn Pro Val Thr Trp Val Lys Thr Ile Gly Ser Thr Thr Ile Ile Asn
625 630 635 640

Leu Ile Leu Ile Leu Val Cys Leu Phe Cys Leu Leu Leu Val Cys Arg
645 650 655

Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp His Arg Glu Arg Ala Met
660 665 670

Met Thr Met Ala Val Leu Ser Lys Arg Lys Gly Gly Asn Val Gly Lys
675 680 685

Ser Lys Arg Asp Gln Ile Val Thr Val Ser Val
690 695

<210> 29

<211> 294

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 29

agttctacaa	tgaacccatc	agagatgcaa	agaaaagcac	ctccgcggag	acggagacat	60
cgcaatcgag	caccgttgac	tcacaagatg	aacaaaatgg	tgacgtcaga	agaacagatg	120
aagttgccat	ccaccaagaa	ggcagagccg	ccaacttggg	cacaactaaa	gaagctgacg	180
cagttagcta	caaaatatct	agagaacaca	aaggtgacac	aaaccccaga	gagtatgctg	240
cttgcagcct	tgatgattgt	atcaatgggtg	gtaagtcctc	ctatgcctgc	agga	294

<210> 30

<211> 57

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 30

tctgcaggtg	tacccaacag	ctccgaagag	acagcgacca	tcgagaacgg	gccatga	57
------------	------------	------------	------------	------------	---------	----

<210> 31

<211> 105

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 31

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg	1	5	10	15
---	---	---	----	----

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr	20	25	30
---	----	----	----

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Gly Pro Pro	35	40	45
---	----	----	----

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu	50	55	60
---	----	----	----

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala	65	70	75	80
---	----	----	----	----

Leu Met Ile Val Ser Met Val Ser Ala Gly Val Pro Asn Ser Ser Glu	85	90	95
---	----	----	----

Glu Thr Ala Thr Ile Glu Asn Gly Pro	100	105
-------------------------------------	-----	-----

<210> 32

Substitute Sequence Listing_USSN 10587032_PP019482.007

<211> 86
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 32
 Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
 1 5 10 15
 His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
 20 25 30
 Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
 35 40 45
 Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
 50 55 60
 Glu Asn Thr Lys Ser Ala Gly Val Pro Asn Ser Ser Glu Glu Thr Ala
 65 70 75 80
 Thr Ile Glu Asn Gly Pro
 85

<210> 33
 <211> 74
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 33
 Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
 1 5 10 15
 Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
 20 25 30
 His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
 35 40 45
 Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
 50 55 60
 Arg His Arg Arg Leu His Phe Val Leu Tyr
 65 70

<210> 34
 <211> 79
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 34
 Met Asn Ser Leu Glu Met Gln Arg Lys Val Trp Arg Trp Arg His Pro
 1 5 10 15
 Asn Arg Leu Ala Ser Leu Gln Val Tyr Pro Ala Ala Pro Lys Arg Gln
 20 25 30
 Gln Pro Ala Arg Met Gly His Ser Asp Asp Gly Gly Phe Val Lys Lys
 35 40 45
 Lys Arg Gly Gly Tyr Val Arg Lys Arg Glu Ile Arg Leu Ser Leu Cys
 50 55 60
 Leu Cys Arg Lys Gly Arg His Lys Lys Leu His Phe Val Leu Tyr

Substitute Sequence Listing_USSN 10587032_PP019482.007

65

70

75

<210>

35

<211>

129

<212>

PRT

<213>

Human endogenous retrovirus, K family (HERV-K)

<400>

35

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Ile Leu Gln Val Tyr Pro Thr Ala Pro Lys Arg
65 70 75 80

Gln Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu
85 90 95

Lys Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr
100 105 110

Cys Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu
115 120 125

Tyr

<210>

36

<211>

125

<212>

PRT

<213>

Human endogenous retrovirus, K family (HERV-K)

<400>

36

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser
65 70 75 80

Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly
85 90 95

Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val
100 105 110

Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu Tyr

Substitute Sequence Listing_USSN 10587032_PP019482.007

115

120

125

<210>

37

<211>

144

<212>

PRT

<213>

Human endogenous retrovirus, K family (HERV-K)

<400>

37

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65 70 75 80

Leu Met Ile Val Ser Met Val Val Tyr Pro Thr Ala Pro Lys Arg Gln
85 90 95

Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys
100 105 110

Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys
115 120 125

Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu Tyr
130 135 140

<210>

38

<211>

74

<212>

PRT

<213>

Human endogenous retrovirus, K family (HERV-K)

<400>

38

Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
1 5 10 15

Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
20 25 30

His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
35 40 45

Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
50 55 60

Arg His Arg Arg Leu His Phe Val Leu Tyr
65 70

<210>

39

<211>

74

<212>

PRT

<213>

Human endogenous retrovirus, K family (HERV-K)

<400>

39

Substitute Sequence Listing_USSN 10587032_PP019482.007

Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
 1 5 10 15
 Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
 20 25 30
 His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
 35 40 45
 Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
 50 55 60
 Arg His Arg Arg Leu His Phe Val Leu Tyr
 65 70

<210> 40
 <211> 44
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 40
 Met Glu Tyr Lys Asn Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Gly
 1 5 10 15
 Asp Ala Lys Lys Arg Ala Ser Thr Glu Met Ser Ala Gly Val Pro Asn
 20 25 30
 Ser Ser Glu Glu Thr Ala Thr Ile Glu Asn Gly Pro
 35 40

<210> 41
 <211> 74
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 41
 Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
 1 5 10 15
 Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
 20 25 30
 His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
 35 40 45
 Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
 50 55 60
 Arg His Arg Arg Leu His Phe Val Leu Tyr
 65 70

<210> 42
 <211> 86
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 42
 Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
 1 5 10 15
 His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
 20 25 30

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
 35 40 45
 Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
 50 55 60
 Glu Asn Thr Lys Ser Ala Gly Val Pro Asn Ser Ser Glu Glu Thr Ala
 65 70 75 80
 Thr Ile Glu Asn Gly Pro
 85

<210> 43
 <211> 105
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 43
 Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
 1 5 10 15
 His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
 20 25 30
 Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
 35 40 45
 Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
 50 55 60
 Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
 65 70 75 80
 Leu Met Ile Val Ser Met Val Ser Ala Gly Val Pro Asn Ser Ser Glu
 85 90 95
 Glu Thr Ala Thr Ile Glu Asn Gly Pro
 100 105

<210> 44
 <211> 127
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 44
 Met Val Thr Pro Val Thr Trp Met Asp Asn Pro Ile Glu Val Tyr Val
 1 5 10 15
 Asn Asp Ser Val Trp Val Pro Gly Pro Thr Asp Asp Arg Cys Pro Ala
 20 25 30
 Lys Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Val Tyr Arg
 35 40 45
 Tyr Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala
 50 55 60
 Val Gln Asn Cys Leu Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg
 65 70 75 80
 Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys
 85 90 95

Substitute Sequence Listing_USSN 10587032_PP019482.007

Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val
100 105 110

Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu Tyr
115 120 125

<210> 45

<211> 105

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 45

Met Val Thr Pro Val Thr Trp Met Asp Asn Pro Ile Glu Val Tyr Val
1 5 10 15

Asn Asp Ser Glu Trp Val Pro Gly Pro Thr Asp Asp Arg Cys Pro Ala
20 25 30

Lys Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Gly Leu Gln
35 40 45

Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly His
50 55 60

Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly Glu
65 70 75 80

Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser Arg
85 90 95

His Arg Arg Leu His Phe Val Met Cys
100 105

<210> 46

<211> 79

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 46

Met Asn Ser Leu Glu Met Gln Arg Lys Val Trp Arg Trp Arg His Pro
1 5 10 15

Asn Arg Leu Ala Ser Leu Gln Val Tyr Pro Ala Ala Pro Lys Arg Gln
20 25 30

Gln Pro Ala Arg Met Gly His Ser Asp Asp Gly Gly Phe Val Lys Lys
35 40 45

Lys Arg Gly Gly Tyr Val Arg Lys Arg Glu Ile Arg Leu Ser Leu Cys
50 55 60

Leu Cys Arg Lys Gly Arg His Lys Lys Leu His Phe Asp Leu Tyr
65 70 75

<210> 47

<211> 214

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 47

Met Asn Ser Leu Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

Substitute Sequence Listing_USSN 10587032_PP019482.007

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Ser Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65 70 75 80

Leu Met Ile Val Ser Met Val Val Ser Leu Pro Met Pro Ala Gly Ala
85 90 95

Ala Ala Ala Asn Tyr Thr Tyr Trp Ala Tyr Val Pro Phe Pro Pro Leu
100 105 110

Ile Arg Ala Val Thr Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn
115 120 125

Asp Ser Val Trp Val Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys
130 135 140

Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr
145 150 155 160

Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val
165 170 175

Gln Asn Trp Leu Val Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe
180 185 190

Thr Tyr His Met Ser Ala Gly Val Pro Asn Ser Ser Glu Glu Thr Ala
195 200 205

Thr Ile Glu Asn Gly Pro
210

<210> 48

<211> 129

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 48

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Thr Leu Gln Val Tyr Pro Thr Ala Pro Lys Arg
65 70 75 80

Gln Arg Pro Ser Arg Thr Gly His Asp Asp Gly Gly Phe Val Glu
85 90 95

Substitute Sequence Listing_USSN 10587032_PP019482.007

Lys Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr
100 105 110

Cys Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Met
115 120 125

Tyr

<210> 49
<211> 125
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 49
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser
65 70 75 80

Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly
85 90 95

Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val
100 105 110

Glu Arg Ser Arg His Arg Arg Leu His Phe Val Met Tyr
115 120 125

<210> 50
<211> 145
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<220>
<221> SITE
<222> 64
<223> Xaa is any amino acid

<400> 50
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Xaa
50 55 60

Leu Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala

Substitute Sequence Listing_USSN 10587032_PP019482.007

65	70	75	80
Ala Leu Met Ile Val Ser Met Val Val Tyr Pro Thr Ala Pro Lys Arg			
	85	90	95
Gln Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu			
	100	105	110
Lys Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr			
	115	120	125
Cys Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Met			
	130	135	140

Tyr
145

<210> 51
<211> 4657
<212> DNA
<213> Artificial Sequence

<220>
<223> pCMVKm2.CORFopt HML-2 vector

<400> 51						
gccgcggaat	ttcgactcta	ggccattgca	tacgtttgat	ctatatcata	atatgtacat	60
ttatattggc	tcatgtccaa	tatgaccgcc	atgttgacat	tgattattga	ctagttatta	120
atagtaatca	attacggggt	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggcccg	ctggctgacc	gcccacgac	ccccgccc	tgacgtcaat	240
aatgacgtat	gttcccatag	taacgccaat	agggactttc	cattgacgtc	aatgggtgga	300
gtatttacgg	taaactgccc	acttggcagt	acatcaagt	tatcatatgc	caagtccgcc	360
ccctattgac	gtcaatgacg	gtaaatggcc	cgcctggcat	tatgcccagt	acatgacctt	420
acgggacttt	cctacttggc	agtacatcta	cgtattagtc	atcgctatta	ccatggtgat	480
gcggttttgg	cagtacacca	atgggcgtgg	atagcggttt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtca	atgggagttt	gttttggcac	caaatcaac	gggactttcc	600
aaaatgtcgt	aataacccc	ccccgttgac	gcaaatgggc	ggtaggcgtg	tacggtggga	660
ggtctatata	agcagagctc	gtttagtga	ccgtcagatc	gcctggagac	gccatccacg	720
ctgttttgac	ctccatgaa	gacaccggga	ccgatccagc	ctccgcggcc	gggaacggtg	780
cattggaacg	cggattcccc	gtgccaagag	tgacgtaagt	accgcctata	gactctatag	840
gcacacccct	ttggctctta	tgcatgctat	actgtttttg	gcttggggcc	tatacacccc	900
cgcttcctta	tgctataggt	gatggtatag	cttagcctat	aggtgtgggt	tattgaccat	960
tattgaccac	tccccatttg	gtgacgatac	tttccattac	taatccataa	catggctctt	1020
tgccacaact	atctctattg	gctatatgcc	aatactctgt	ccttcagaga	ctgacacgga	1080
ctctgtattt	ttacaggatg	gggtcccatt	tattattttac	aaattcacat	atacaacaac	1140
gccgtccccc	gtgcccgcag	tttttattaa	acatagcgtg	ggatctccac	gcgaatctcg	1200
ggtacgtgtt	ccggacatgg	gctcttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
tggtcccatg	cctccagcgg	ctcatggtcg	ctcggcagct	ccttgctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccagt	tgccgcacaa	ggccgtggcg	1380
gtagggatat	tgtctgaaaa	tgagctcgga	gattgggctc	gcaccgctga	cgcatatgga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagctgag	ttgttgatt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgctgtta	acggtggagg	gcagtgtagt	ctgagcagta	1560
ctcgttgctg	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttcctt	1620
tccatgggtc	ttttctgcag	tcaccgtcgt	cgacgccacc	atgaaccca	gcgagatgca	1680
gcgcaaggcc	cccccccgcc	gccgccgcca	ccgcaaccgc	gccccctga	cccacaagat	1740
gaacaagatg	gtgaccagcg	aggagcagat	gaagctgccc	agcaccaaga	aggccgagcc	1800
ccccacctgg	gcccagctga	agaagctgac	ccagctggcc	accaagtacc	tggagaacac	1860
caaggtgacc	cagacccccg	agagcatgct	gctggccgcc	ctgatgatcg	tgagcatggt	1920
gagcgccggc	gtgccaaca	gcagcgagga	gaccgccacc	atcgagaacg	gccccgctta	1980
aagaattcag	actcgagcaa	gtctagaaag	ccatggatat	cgatccact	acgcgttaga	2040
gctcgtgat	cagcctcgac	tgtgccttct	agttgccagc	catctgttgt	ttgcccctcc	2100
cccgtgcctt	ccttgaccct	ggaaggtgcc	actcccactg	tcctttccta	ataaaatgag	2160
gaaattgcat	cgcatgtct	gagtagggtg	cattctattc	tgggggggtg	ggtggggcag	2220

Substitute Sequence Listing_USSN 10587032_PP019482.007

gacagcaagg	gggaggattg	ggaagacaat	agcagggggg	tgggcgaaga	actccagcat	2280
gagatccccg	cgctggagga	tcattccagcc	ggcgtcccg	aaaacgattc	cgaagcccaa	2340
cctttcatag	aaggcgccg	tggaatcgaa	atctcgtgat	ggcaggttgg	gcgtcgcttg	2400
gtcggtcatt	tcgaacccca	gagtcccgt	cagaagaact	cgtaagaag	gcgatagaag	2460
gcgatgcgt	gcgaatcggg	agcggcgata	ccgtaaagca	cgaggaagcg	gtcagcccat	2520
tcgccgcca	gctcttcagc	aatatcacgg	gtagccaacg	ctatgtcctg	atagcgggtcc	2580
gccacacca	gccggccaca	gtcgatgaat	ccagaaaagc	ggccattttc	caccatgata	2640
ttcggcaagc	aggcatcgcc	atgggtcacg	acagatcct	cgccgtcggg	catgcgcgcc	2700
ttgagcctgg	cgaacagttc	ggctggcgcg	agcccctgat	gctcttcgtc	cagatcatcc	2760
tgatcgacaa	gaccggcttc	catccgagta	cgtgctcgct	cgatgcgatg	tttcgcttgg	2820
tggtcgaatg	ggcaggtagc	cggatcaagc	gtatgcagcc	gccgcattgc	atcagccatg	2880
atggatactt	tctcggcagg	agcaagggtga	gatgacagga	gatcctgccc	cggcacttcg	2940
cccaatagca	gccagtcctt	tcccgttca	gtgacaacgt	cgagcacagc	tgcgcaagga	3000
acgcccgtcg	tggccagcca	cgatagccgc	gctgcctcgt	cctgcagttc	attcagggca	3060
ccggacaggt	cggtccttgac	aaaaagaacc	ggcgccccct	cgctgacag	ccggaacacg	3120
gcggcatcag	agcagccgat	tgtctgttgt	gcccagtcac	agccgaatag	cctctccacc	3180
caagcggccg	gagaacctgc	gtgcaatcca	tcttgttcaa	tcatgcgaaa	cgatcctcat	3240
cctgtctctt	gatcagatct	tgatccccctg	cgccatcaga	tccttggcgg	caagaaagcc	3300
atccagtitta	ctttgcaggg	cttcccaacc	ttaccagagg	gcgccccagc	tggcaattcc	3360
ggttcgcttg	ctgtccataa	aaccgcccag	tctagtatc	gccatgtaag	cccactgcaa	3420
gctacctgct	ttctctttgc	gcttgcgttt	tcccttgtcc	agatagccca	gtagctgaca	3480
ttcatccggg	gtcagcaccg	tttctgcgga	ctggctttct	acgtgttccg	cttcctttag	3540
cagcccttgc	gccctgagtg	cttgccggcag	cgtgaagcta	attcatggtt	aaatttttgt	3600
taaatcagct	catttttttaa	ccaataggcc	gaaatcggca	aaatccctta	taaatcaaaa	3660
gaatagcccg	agatagggtt	gagtgttgtt	ccagtttggg	acaagagtcc	actattaaag	3720
aacgtggagt	ccaacgtcaa	agggcgaaaa	accgtctatc	agggcgatgg	ccggatcagc	3780
ttatgcggtg	tgaatatccg	cacagatgcg	taaggagaaa	ataccgcac	aggcgctctt	3840
ccgcttctct	gctcactgac	tcgctgcgct	cggtcgttcg	gctgcggcga	gcggtatcag	3900
ctcactcaaa	ggcggtataa	cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	3960
tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	ggccgcgttg	ctggcgtttt	4020
tccataggct	ccgccccctt	gacgagcatc	acaaaaatcg	acgctcaagt	cagaggtggc	4080
gaaacccgag	aggactataa	agataccagg	cgtttcccc	tggaaagtcc	ctcgtgcgct	4140
ctcctgttcc	gaccctggcg	cttaccggat	accgttccgc	ctttctccct	tcgggaagcg	4200
tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	ggtgtaggtc	gttcgctcca	4260
agctgggctg	tgtgcacgaa	cccccggttc	agcccgaccg	ctgcgcctta	tccggttaact	4320
atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	actggcagca	gccactggta	4380
acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	gttcttgaag	tgggtggccta	4440
actacggcta	cactagaagg	acagtatttg	gtatctgcgc	tctgctgaag	ccagttacct	4500
tcggaaaaag	agttggtagc	tcttgatccg	gcaaacaaac	caccgctggt	agcggtggtt	4560
ttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	atctcaagaa	gatcctttga	4620
tcttttctac	tgaacggtga	tccccaccgg	aattgcg			4657

<210> 52
 <211> 4774
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pCMVKm2.pCAP5opt HML-2 vector

<400> 52						
gccgcggaat	ttcgactcta	ggccattgca	tacgtttgtat	ctatatcata	atatgtacat	60
ttatattggc	tcattgtccaa	tatgaccgcc	atgttgacat	tgattattga	ctagttatta	120
atagtaatca	attacggggg	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggccccg	ctggctgacc	gcccacgac	ccccgccc	tgacgtcaat	240
aatgacgtat	gttcccatag	taacgccaat	agggactttc	cattgacgtc	aatgggtgga	300
gtatttcagg	taaaactgcc	acttggcagt	acatcaagt	tatcatatgc	caagtccgcc	360
ccctattgag	gtcaatgcag	gtaaatggcc	cgctggcat	tatgcccagt	acatgacctt	420
acgggacttt	cctacttggc	agtacatcta	cgtattagtc	atcgctatta	ccatggtgat	480
gcggttttgg	cagtacacca	atgggcgtgg	atagcggttt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtca	atgggagttt	gttttggcac	caaaatcaac	gggactttcc	600
aaaatgtcgt	aataaccccc	ccccgttgac	gcaaatgggc	ggtaggcgtg	tacgggtggga	660
ggtctatata	agcagagctc	gtttagtga	ccgtcagatc	gcctggagac	gccatccacg	720

Substitute Sequence Listing_USSN 10587032_PP019482.007

ctgtttttgac	ctccatagaa	gacaccggga	ccgatccagc	ctccgcggcc	gggaacggtg	780
catttggaacg	cggattcccc	gtgccaaagag	tgacgtaagt	accgcctata	gactctatag	840
gcacacccct	ttggctctta	tgcatgctat	actgtttttg	gcttggggcc	tatacacccc	900
cgcttcctta	tgctataggt	gatggtatag	cttagcctat	aggtgtgggt	tattgaccat	960
tattgaccac	tccccatttg	gtgacgatac	tttcattac	taatccataa	catggctctt	1020
tgccacaact	atctctattg	gctatatgcc	aatactctgt	ccttcagaga	ctgacacgga	1080
ctctgtattt	ttacaggatg	gggtcccatt	tattatttac	aaattcacat	atacaacaac	1140
gccgtcccc	gtgcccgcag	tttttattaa	acatagcgtg	ggatctccac	gcgaatctcg	1200
ggtacgtggt	ccggacatgg	gctcttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
tggtcccatg	cctccagcgg	ctcatggtcg	ctcggcagct	ccttgctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccagtg	tgccgcacaa	ggccgtggcg	1380
gtagggatg	tgtctgaaaa	tgagctcgga	gattgggctc	gcaccgctga	cgcatagga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagctgag	ttgttgatt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgctgtta	acggtggagg	gcagtgtagt	ctgagcagta	1560
ctcgttgctg	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttcctt	1620
tccatgggtc	ttttctgcag	tcaccgtcgt	cgacgccacc	atgaaccca	gcgagatgca	1680
gcgcaaggcc	cccccccgcc	gccgcgcgca	ccgcaaccgc	gccccctga	cccacaagat	1740
gaacaagatg	gtgaccagcg	aggagcagat	gaagctgccc	agcaccaaga	aggccgagcc	1800
ccccacctgg	gcccagctga	agaagctgac	ccagctggcc	accaagtacc	tggagaacac	1860
caaggtgacc	cagacccccg	agagcatgct	gctggccgcc	ctgatgatcg	tgagcatggt	1920
ggtgtacccc	accgccccca	agcgccagcg	ccccagccgc	accggccacg	acgacgacgg	1980
cggtctcgtg	gagaagaagc	gcggcaagtg	cggcgagaag	caggagcgca	gcgactgcta	2040
ctgctgtgct	gtggagcgca	gccgccaccg	ccgcctgcac	ttcgtgctgt	acgcttaaag	2100
aattcagact	cgagcaagtc	tagaaagcca	tggatatcgg	atccactacg	cgtagagct	2160
cgctgatcag	cctcgactgt	gccttctagt	tgccagccat	ctgttgtttg	cccccccc	2220
gtgccttctt	tgaccctgga	aggtgccatt	cccactgtcc	tttcctaata	aaatgaggaa	2280
attgcatcgc	attgtctgag	taggtgtcat	tctattctgg	ggggtggggg	ggggcaggac	2340
agcaaggggg	aggattggga	agacaatagc	aggggggtgg	gcgaagaact	ccagcatgag	2400
atccccgcgc	tggaggatca	tccagccggc	gtcccggaaa	acgattccga	agcccaacct	2460
ttcatagaag	gcggcggtgg	aatcgaaatc	tcgtgatggc	aggttgggcg	tcgcttggtc	2520
ggtcatttcg	aaccccagag	tcccgtctag	aagaactcgt	caagaaggcg	atagaaggcg	2580
atgcgtctcg	aatcgggagc	ggcgataccg	taaagcaca	ggaagcggtc	agcccattcg	2640
ccgccaagct	cttcagcaat	atcacgggta	gccaacgcta	tgtcctgata	gcggctccgc	2700
acaccagacc	ggccacagtc	gatgaatcca	gaaaagcggc	cattttccac	catgatattc	2760
ggcaagcagg	catcgccatg	ggtcacgacg	agatcctcgc	cgtcgggcat	gcgcgccttg	2820
agcctggcga	acagttcggc	tggcgcgagc	ccctgatgct	cttcgtccag	atcatcctga	2880
tcgacaagac	cggcttccat	ccgagtacgt	gctcgtcga	tgcatgttt	cgcttggtgg	2940
tcgaatgggc	aggtagccgg	atcaagcgta	tgacgcgcgc	gcattgcata	agccatgatg	3000
gatactttct	cggcaggagc	aaggtgagat	gacaggagat	cctgccccgg	cacttcgccc	3060
aatagcagcc	agtcccttcc	cgcttcagtg	acaacgtcga	gcacagctgc	gcaagggaacg	3120
cccgtcgtgg	ccagccacga	tagccgcgct	gcctcgtcct	gcagttcatt	cagggcaccg	3180
gacaggtcgg	tcttgacaaa	aagaaccggg	cgcccctgcg	ctgacagccg	gaacacggcg	3240
gcatcagagc	agccgattgt	ctgttggtgc	cagtcatagc	cgaatagcct	ctccacccaa	3300
gcggccggag	aacctgcgtg	caatccatct	tgttcaatca	tgcgaaacga	tcctcatcct	3360
gtctcttgat	cagatcttga	tcccctgcgc	catcagatcc	ttggcggcaa	gaaagccatc	3420
cagtttactt	tgcagggtct	cccaacctta	ccagagggcg	ccccagctgg	caattccggt	3480
tcgcttgctg	tccataaaac	cgcccagtct	agctatcgcc	atgtaagccc	actgcaagct	3540
acctgctttc	tctttgcgct	tgctgttttc	cttgctcaga	tagcccagta	gctgacattc	3600
atccgggggtc	agcaccgttt	ctgcggactg	gctttctacg	tgttccgctt	ccttttagcag	3660
cccttgcgcc	ctgagtgcct	gcggcagcgt	gaagctaatt	catggttaaa	tttttgtaa	3720
atcagctcat	tttttaacca	ataggccgaa	atcggcaaaa	tcccttataa	atcaaaagaa	3780
tagcccagaga	taggggttag	tgttggtcca	gtttggaaca	agagtccact	attaaagaac	3840
gtggactcca	acgtcaaagg	gcgaaaaacc	gtctatcagg	gcgatggccg	gatcagctta	3900
tgcggtgtga	aataccgcac	agatgcgtaa	ggagaaaata	ccgcatcagg	cgctcttcg	3960
cttctctgct	cactgactcg	ctgcgctcgg	tcgctcggct	gcggcgagcg	gtatcagctc	4020
actcaaaggc	ggtaatacgg	ttatccacag	aatcagggga	taacgcagga	aagaacatgt	4080
gagcaaaagg	ccagcaaaag	gccaggaacc	gtaaaaaagg	cgcgttgctg	gcgtttttcc	4140
ataggtctcg	ccccctgac	gagcatcaca	aaaatcgacg	ctcaagtcag	aggtggcgaa	4200
acccgacagg	actataaaga	taccaggcgt	ttccccctgg	aagctccctc	gtgcgctctc	4260
ctgttccgac	cctgccgctt	accggatacc	tgctccgctt	tctcccttcg	ggaagcgtgg	4320
cgctttctca	tagctcacgc	tgtaggtatc	tcagttcggg	gtaggtcggt	cgctccaagc	4380
tgggctgtgt	gcacgaaccc	cccgttcagc	ccgaccgctg	cgcttatcc	ggtaactatc	4440
gtcttgagtc	caacccggtg	agacacgact	tatcgccact	ggcagcagcc	actggttaaca	4500

Substitute Sequence Listing_USSN 10587032_PP019482.007

ggattagcag	agcgaggtat	gtaggcgggtg	ctacagagtt	cttgaagtgg	tggcctaact	4560
acggctacac	tagaaggaca	gtatttggtg	tctgcgctct	gctgaagcca	gttaccttcg	4620
gaaaaagagt	tggtagctct	tgatccggca	aacaaaccac	cgctggtagc	ggtgggtttt	4680
ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaaggatc	tcaagaagat	cctttgatct	4740
tttctactga	acggtgatcc	ccaccggaat	tgcg			4774

<210> 53
 <211> 6483
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pCMVKm2.gag wt PCAV vector

<400> 53						
gccgcggaat	ttcgactcta	ggccattgca	tacgtttgtat	ctatatcata	atatgtacat	60
ttatattggc	tcatgtccaa	tatgaccgcc	atgttgacat	tgattattga	ctagttatta	120
atagtaatca	attacggggg	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggccccg	ctggctgacc	gcccacacac	ccccgccc	tgacgtcaat	240
aatgacgtat	gttcccatag	taacgccaat	agggactttc	cattgacgtc	aatgggtgga	300
gtattttacg	taaaactggc	acttggcagt	acatcaagtg	tatcatatgc	caagtccgac	360
ccctattgac	gtcaatgacg	gtaaatggcc	cgcctggcat	tatgcccagt	acatgacctt	420
acgggacttt	cctacttggc	agtacatcta	cgtattagtc	atcgctatta	ccatggtgat	480
gcggttttgg	cagtacacca	atgggcgtgg	atagcggttt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtca	atgggagttt	gttttggcac	caaaatcaac	gggactttcc	600
aaaatgtcgt	aataaaccgg	ccccgttgac	gcaaattggc	ggtaggcgtg	tacggtggga	660
ggtctatata	agcagagctc	gtttagtga	ccgtcagatc	gcctggagac	gccatccacg	720
ctgttttgac	ctccatagaa	gacaccggga	ccgatccagc	ctccgcgcc	gggaacggtg	780
cattggaacg	cggattcccc	gtgccaagag	tgacgtaagt	accgcctata	gactctatag	840
gcacacccct	ttggctctta	tgcatgctat	actgtttttg	gcttggggcc	tatacacccc	900
cgcttcctta	tgctataggt	gatggtatag	cttagcctat	aggtgtgggt	tattgaccat	960
tattgaccac	tcccctattg	gtgacgatac	tttccattac	taatccataa	catggctctt	1020
tgccacaact	atctctattg	gcttatatgc	aataactctg	ccttcagaga	ctgacacgga	1080
ctctgtattt	ttacaggatg	gggtcccatt	tattattttac	aaattcacat	atacaacaac	1140
gccgtccccc	gtgcccgcag	tttttattaa	acatagcgtg	ggatctccac	gcgaatctcg	1200
ggtacgtgtt	ccggacatgg	gctcttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
tggtcccatg	cctccagcgg	ctcatggctg	ctcggcagct	ccttgctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccagtg	tgccgcacaa	ggccgtggcg	1380
gtagggtatg	tgtctgaaaa	tgagtctgca	gatttggctc	gcaccgtga	cgcatatgga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagctgag	ttgtttgtatt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgctgtta	acggtggagg	gcagtgtagt	ctgagcagta	1560
ctcgttgctg	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttcctt	1620
tccatgggtc	ttttctgcag	tcaccgtcgt	cgacgccacc	atggggcaaa	ctgaaagtaa	1680
atatgcctct	tatctcagct	ttattaaaat	tcttttaaga	agagggggag	ttagagcttc	1740
tacagaaact	ctaattacgc	tatttcaaac	aataagaaca	ttctgcccac	ggtttccaga	1800
acagggaaact	ttagatctaa	aagattggga	aaaaattggc	aaagaattaa	aacaagcaaa	1860
taggggaagg	aaaatcatcc	cacttacagt	atggaatgat	tggggcatta	ttaaagcaac	1920
tttagaacca	tttcaaacag	gagaagatat	tgtttcagtt	tctgatgcc	ctaaaagctg	1980
tgtaacagat	tgtgaagaag	aggcagggac	agaatcccag	caagggaacg	aaagttcaca	2040
ttgtaaatat	gtagcagagt	ctgtaatggc	tcagtcaacg	caaaatgttg	actacagtca	2100
attacaggag	ataatatacc	ctgaatcac	aaaattgggg	gaaggaggtc	cagaatcatt	2160
ggggccatca	gagcctaacc	cacgatcgcc	atcaactcct	cctcccgtgg	ttcagatgcc	2220
tgtaacatta	caacctcaaa	cgcagggttag	acaagcacia	accccaagag	aaaatcaagt	2280
agaaagggac	agagtctcta	tcccggcaat	gccaactcag	atacagtatc	cacaatatca	2340
gccggtagaa	aataagaccc	aaccgctgg	agtttatcaa	taccggctgc	caaccgagct	2400
tcagtatcgg	cctccttcag	aggttcaata	cagacctcaa	gcggtgtgtc	ctgtgcaaaa	2460
tagcacggca	ccataccagc	aaccacacg	gatggcgtct	aattcaccag	caacacagga	2520
cgcgcgctg	tatcttcagc	cgccactgt	gagacttaat	cctacagcat	cacgtagtgg	2580
acaggggtgt	gcactgcatg	cagtcattga	tgaagccaga	aaacagggcg	atcttgaggc	2640
atggcggttc	ctggtaattt	tacaactgg	acaggccggg	gaagagactc	aagtaggagc	2700
gcctgcccga	gctgagacta	gatgtgaacc	tttcaccatg	aaaatgttaa	aagatataaa	2760
ggaaggagtt	aaacaatatg	gatccaactc	cccttatata	agaacattat	tagattccat	2820
tgctcatgga	aatagactta	ctccttatga	ctgggaattt	ttggccaaat	cttccctttc	2880

Substitute Sequence Listing_USSN 10587032_PP019482.007

atcctctcag	tatctacagt	ttaaaacctg	gtggattgat	ggagtacaag	aacaggtacg	2940
aaaaaatcag	gctactaagc	ccactgttaa	tatagacgca	gaccaattgt	taggaacagg	3000
tccaaattgg	agcaccatta	accaacaatc	agtgatgcag	aatgaggcta	ttgaacaagt	3060
aagggctatt	tgcctcaggg	cctggggaaa	aattcaggac	ccaggaacag	ctttccctat	3120
taattcaatt	agacaaggct	ctaaagagcc	atatictgac	tttgtggcaa	gattacaaga	3180
tgctgctcaa	aagtctatta	cagatgacaa	tgcccgaaaa	gttattgtag	aattaatggc	3240
ctatgaaaat	gcaaatccag	aatgtcagtc	ggccataaag	ccattaaaag	gaaaagttcc	3300
agcaggagtt	gatgttaatta	cagaatatgt	gaaggcttgt	gatgggattg	gaggagctat	3360
gcataaggca	atgctaattg	ctcaagcaat	gagggggctc	actctaggag	gacaagttag	3420
aacatttggg	aaaaaatgtt	ataattgtgg	tcaaactcgg	catctgaaaa	ggagttgccc	3480
agtcttaaat	aaacagaata	taataaatca	agctattaca	gcaaaaaata	aaaagccatc	3540
tggcctgtgt	ccaaaatgtg	gaaaaggaaa	acattggggc	aatcaatgtc	attctaaatt	3600
tgataaggat	gggcaaccat	tgtcgggaaa	caggaagagg	ggccagcctc	agggccccca	3660
acaaactggg	gatttcccg	ttcaactggt	tgttcttcag	ggtttttcaag	gacaacaacc	3720
cctacagaaa	ataccaccac	ttcagggaat	cagccaatta	caacaatcca	acagctgtcc	3780
cgcgccacag	caggcagcac	cgcatgaaga	attcagactc	gagcaagtct	agaaagccat	3840
ggatatcgga	tccactacgc	gttagagctc	gctgatcagc	ctcgaactgt	ccttctagtt	3900
gccagccatc	tgttgtttgc	ccctcccccg	tgccttcctt	gaccctggaa	ggtgccactc	3960
ccactgtcct	ttcctaataa	aatgaggaaa	ttgcatcgca	ttgtctgagt	aggtgtcatt	4020
ctattctggg	gggtgggggt	gggcaggaca	gcaaggggga	ggattgggaa	gacaatagca	4080
gggtgggtgg	cgaaagaactc	cagcatgaga	tccccgcgct	ggaggatcat	ccagccggcg	4140
tcccggaaaa	cgattccgaa	gcccacactt	tcatagaagg	cggcgggtga	atcgaaatct	4200
cgatgatggc	ggttgggctg	cgcttggtcg	gtcatttcga	accccagagt	cccgtctaga	4260
agaactcgtc	aagaaggcga	tagaaggcga	tgcgctgcga	atcgggagcg	gcgataccgt	4320
aaagcacgag	gaagcggcta	gcccattcgc	cgccaagctc	ttcagcaata	tcacgggtag	4380
ccaacgctat	gtcctgatag	cggtccgcca	cacccagcgc	gccacagtcg	atgaatccag	4440
aaaagcggcc	atgttccacc	atgatattcg	gcaagcaggc	atcgccatgg	gtcacgcaga	4500
gatcctcgcc	gtcgggcatg	cgcgcttga	gcctggcgaa	cagttcggct	ggcgcgagcc	4560
cctgatgtc	ttcgtccaga	tcacctctgat	cgacaagacc	ggcttccatc	cgagtacgtg	4620
ctcgtctgat	gcgatgtttc	gcttgggtgt	cgaatgggca	ggtagccgga	tcaagcgtat	4680
gcagccgccc	cattgcatca	gccatgatgg	atactttctc	ggcaggagca	aggtgagatg	4740
acaggagatc	ctgccccggc	acttcgcccc	atagcagcca	gtcccttccc	gcttcagtga	4800
caacgtctgag	cacagctgcg	caaggaaacg	ccgtcgtggc	cagccacgat	agccgcgctg	4860
cctcgtcctg	cagttcattc	agggcaccgg	acaggtcggg	cttgacaaaa	agaaccgggc	4920
gcccctgcgc	tgacagccgg	aacacggcgg	catcagagca	gccgattgtc	tggtgtgccc	4980
agtcatagcc	gaatagcctc	tccacccaag	cggccggaga	acctgcgtgc	aatccatctt	5040
gttcaatcat	gcgaaacgat	cctcatcctg	tctcttgatc	agatcttgat	cccctgcgcc	5100
atcagatcct	tggcggcaag	aaagccatcc	agtttacttt	gcagggtctc	ccaaccttac	5160
cagagggcgc	cccagctggc	aattccggtt	cgttgctgtg	ccataaaaac	gcccagtcta	5220
gctatcgcca	tgtaagccca	ctgcaagcta	cctgtctttc	ctttgcgctt	gcgttttccc	5280
ttgtccagat	agcccagtag	ctgacattca	tccggggctc	gcaccgtttc	tgccgactgg	5340
ctttctacgt	gttccgcttc	ctttagcagc	ccttgccgcc	tgagtgtgtg	cggcagcgtg	5400
aagctaattc	atgggttaaat	ttttgttaaa	tcagctcatt	ttttaaccaa	taggccgaaa	5460
tcggcaaaa	cccttataaa	tcaaaaagaat	agcccagat	agggttgagt	gttgttccag	5520
tttggaaaca	gagtcacta	ttaaagaacg	tggtactcaa	cgtcaaaggg	cgaaaaaccg	5580
tctatcaggg	cgatggccgg	atcagcttat	gcggtgtgaa	ataccgcaca	gatgcgtaag	5640
gagaaaaatac	cgcatcaggc	gctcttcgcg	ttcctcgtc	actgactcgc	tgcgctcgg	5700
cgttcggctg	cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacgg	tatccacaga	5760
atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	5820
taaaaaggcc	gcgttgctgg	cgtttttcca	taggtctcgc	ccccctgacg	agcatcacia	5880
aaatcgacgc	tcaagtcaga	ggtggcgaaa	cccagacagg	ctataaagat	accaggcggt	5940
tccccctgga	agctccctcg	tgcgctctcc	tgttccgacc	ctgccgctta	ccggatacct	6000
gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	6060
cagttcgggtg	taggtcggtc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	6120
cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtcc	aacccggtaa	gacacgactt	6180
atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgaggatg	taggcgggtg	6240
tacagagttc	ttgaagtgg	ggcctaacta	cggctacagt	agaaggacag	tatttggtat	6300
ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	6360
acaaaccacc	gctggtagcg	gtggtttttt	tgtttgcaag	cagcagatta	cgcgagaaa	6420
aaaaggatct	caagaagatc	ctttgatctt	ttctactgaa	cggtgatccc	caccggaatt	6480
gcg.						6483

Substitute Sequence Listing_USSN 10587032_PP019482.007

<211> 6340

<212> DNA

<213> Artificial Sequence

<220>

<223> pCMVKm2.gagopt HML-2 vector

<400> 54

gccgcggaat	ttcgactcta	ggccattgca	tacgtttgat	ctatatcata	atatgtacat	60
ttatattggc	tcatgtccaa	tatgaccgcc	atgttgacat	tgattattga	ctagttatta	120
atagtaatca	attacggggt	cattagtcca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggcccgc	ctggctgacc	gcccacacgac	ccccgcccac	tgacgtcaat	240
aatgacgat	gttcccatag	taacgccaat	agggactttc	cattgacgct	aatgggtgga	300
gtatttacgg	taaaactgccc	acttggcagt	acatcaagt	tatcatatgc	caagtccgcc	360
ccctattgac	gtcaatgacg	gtaaaatggcc	cgcttggcat	tatgcccagt	acatgacctt	420
acgggacttt	cctacttggc	agtacatcta	cgtattagtc	atcgctatta	ccatgggtgat	480
gcggttttgg	cagtacacca	atgggcgtgg	atagcggttt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtca	atggggagttt	gttttggcac	caaaatcaac	gggactttcc	600
aaaatgtcgt	aataacccc	ccccgttgac	gcaaattgggc	ggtaggcgtg	tacgggtgga	660
ggtctatata	agcagagctc	gttttagtgaa	ccgtcagatc	gcctggagac	gccatccacg	720
ctgttttgac	ctccatagaa	gacaccggga	ccgatccagc	ctccgcggcc	gggaacgggtg	780
cattggaacg	cggattcccc	gtgccaagag	tgacgtaagt	accgcctata	gactctatag	840
gcacacccct	ttggctctta	tgcattgctat	actgtttttg	gcttggggcc	tatacacccc	900
cgcttcctta	tgctataggt	gatggtatag	cttagcctat	aggtgtgggt	tattgaccat	960
tattgaccac	tccccatttg	gtgacgatac	tttccattac	taatccataa	catggctctt	1020
tgccacaact	atctctattg	gctatatgcc	aatactctgt	ccttcagaga	ctgacacgga	1080
ctctgtattt	ttacaggatg	gggtcccatt	tattattttac	aaattcacat	atacaacaac	1140
gccgtccccc	gtgcccgcag	tttttattaa	acatagcgtg	ggatctccac	gcgaatctcg	1200
ggtacgtgtt	ccggacatgg	gctcttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
tggtcccctg	cctccagcgg	ctcatggtcg	ctcggcagct	ccttgctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccagt	tgccgcacaa	ggccgtggcg	1380
gtagggtagt	tgtctgaaaa	tgagctcgga	gattgggctc	gcaccgctga	cgcatatgga	1440
agacttaagg	cagcggcgaga	agaagatgca	ggcagctgag	ttgttgtatt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgctgtta	agcgtggagg	gcagtgtagt	ctgagcagta	1560
ctcgttgctg	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttcctt	1620
tccatgggtc	ttttctgcag	tcaccgtcgt	cgacgccacc	atggggcaga	ccaagagcaa	1680
gatcaagagc	aagtacgcca	gctacctgag	cttcatcaag	atcctgctga	agcgcggcgg	1740
cgtgaagggt	agcaccaaga	acctgatcaa	gctgttccag	atcatcgagc	agttctgccc	1800
ctggttcccc	gagcaggcca	ccctggacct	gaaggactgg	aagcgcacgt	gcaaggagct	1860
gaagcaggcc	ggccgcgaagg	gcaacatcat	ccccctgacc	gtgtggaacg	actgggccat	1920
catcaaggcc	gccctggagc	ccttccagac	cgaggaggac	agcgtgagcg	tgagcgacgc	1980
ccccggcagc	tgcatcatcg	actgcaacga	gaacaccgcg	aagaagagcc	agaaggagac	2040
cgagggcctg	cactgcgagt	acgtggccga	gcccgtgatg	gcccagagca	cccagaacgt	2100
ggactacaac	cagctgcagg	aggtgatcta	ccccgagacc	ctgaagctgg	agggcaaggg	2160
ccccgtcgtg	gtggggccca	gcgagagcaa	gccccgcggc	accagcccc	tgcccgccgg	2220
ccaggtgccc	gtgaccctgc	agccccagaa	gcaggtgaag	gagaacaaga	cccagcccc	2280
cgtggcctac	cagtactggc	ccccgcgcga	gctgcagtac	cgcccccccc	ccgagagcca	2340
gtacggctac	cccggcatgc	ccccgcgcgc	ccagggccgc	gccccctacc	cccagcccc	2400
caccgcgcgc	ctgaacccca	ccgccccccc	cagccgccag	ggcagcaagc	tgacagagat	2460
catcgacaag	agccgcaagg	agggcgacac	cgaggcctgg	cagttccccg	tgaccctgga	2520
gccccatgccc	cccggcgagg	gcgcccagga	gggcgagccc	cccaccgtgg	aggcccgccta	2580
caagagcttc	agcatcaaga	agctgaagga	catgaaggag	ggcgtgaagc	agtacggccc	2640
caacagcccc	tacatgcgca	ccctgctgga	cagcatcgcc	cacggccacc	gcctgatccc	2700
ctacgactgg	gagatcctgg	ccaagagcag	cctgagcccc	agccagttcc	tgcatgtcaa	2760
gacctgggtg	atcgacggcg	tgacaggagca	ggtgcgccgc	aaccgcgccg	ccaaccccc	2820
cgtgaacatc	gacgccgacc	agctgctggg	catcggccag	aactggagca	ccatcagcca	2880
gcaggccctg	atgcagaaacg	aggccatcga	gcaggtgcgc	gccatctgcc	tgccgcctcg	2940
ggagaagatc	caggacccccg	gcagcacctg	ccccagcttc	aacaccgtgc	gccagggcag	3000
caaggagccc	taccccgact	tcgtggccccg	cctgcaggac	gtggcccaga	agagcatcgc	3060
cgacgagaag	gcccgcgaagg	tgatcgtgga	gctgatggcc	tacgagaacg	ccaaccccga	3120
gtgccagagc	gccatcaagc	ccctgaaggg	caaggtgccc	gcccgcagcg	acgtgatcag	3180
cgagtacgtg	aaggcctgcg	acggcatcgg	cggcgccatg	cacaaggcca	tgctgatggc	3240
ccaggccatc	accggcgctgg	tgctgggcgg	ccaggtgcgc	accttcggcc	gcaagtgccta	3300

Substitute Sequence Listing_USSN 10587032_PP019482.007

caactgcggc	cagatcgcc	acctgaagaa	gaactgcccc	gtgctgaaca	agcagaacat	3360
caccatccag	gccaccacca	ccggcccgca	gccccccgac	ctgtgcccc	gctgcaagaa	3420
gggcaagcac	tgggccagcc	agtgccgcag	caagttcgac	aagaacggcc	agcccctgag	3480
cggcaacgag	cagcgcggcc	agccccaggc	ccccccagcag	accggcgcc	tccccatcca	3540
gcccttcgtg	ccccagggct	tccagggcca	gcagcccccc	ctgagccagg	tgttccaggg	3600
catcagccag	ctgccccagt	acaacaactg	ccccccccc	caggccgccc	tgcagcaggc	3660
ttaaagaatt	cagactcgag	caagtctaga	aagccatgga	tatcggatcc	actacgcgtt	3720
agagctcgct	gatcagcctc	gactgtgcct	tctagtggcc	agccatctgt	tgtttgcccc	3780
tccccgtgc	cttccttgac	cctggaaggt	gccactccca	ctgtcctttc	ctaataaaat	3840
gaggaaattg	catcgcattg	tctgagtagg	tgtcattcta	ttctgggggg	tggggtgggg	3900
caggacagca	agggggagga	ttgggaagac	aatagcaggg	gggtgggcga	agaactccag	3960
catgagatcc	ccgcgctgga	ggatcatcca	gccggcgctc	cggaaaacga	ttccgaagcc	4020
caacctttca	tagaaggcgg	cgggtggaatc	gaaatctcgt	gatggcagg	tgggcgtcgc	4080
ttggtcggtg	atttcgaacc	ccagagtcctc	atcagaaga	actcgtcaag	aaggcgatag	4140
aaggcagtc	gctgcgaatc	gggagcggcg	ataccgtaaa	gcacgaggaa	gcggtcagcc	4200
cattcgccgc	caagctcttc	agcaatatca	cgggtagcca	acgctatgtc	ctgatagcgg	4260
tccgccacac	ccagccggcc	acagtcgatg	aatccagaaa	agcggccatt	ttccaccatg	4320
atattcgga	agcaggcatc	gccatgggtc	acgacgagat	cctcgccgct	gggcatgcgc	4380
gccttgagcc	tggcgaacag	ttcggctggc	gcgagcccc	gatgctcttc	gtccagatca	4440
tcctgatcga	caagaccggc	ttccatccga	gtacgtgctc	gctcgatgcg	atgtttcgct	4500
tggttgctga	atgggcagg	agccggatca	agcgtatgca	gccgcccgat	tgcatcagcc	4560
atgatggata	ctttctcggc	aggagcaagg	tgagatgaca	ggagatcctg	ccccggcact	4620
tcgccaata	gcagccagtc	ccttcccgtc	tcagtgaaca	cgtcgagcac	agctgcgcaa	4680
ggaacgccc	tcgtggccag	ccacgatagc	cgcgctgcct	cgctctgcag	ttcattcagg	4740
gcaccggaca	ggtcggctct	gacaaaaaga	accgggcgcc	cctgcgctga	cagccggaac	4800
acggcggcag	cagagcagcc	gattgtctgt	tgtgcccagt	catagccgaa	tagcctctcc	4860
acccaagcgg	ccggagaacc	tgctgtgcat	ccatcttggt	caatcatgcg	aaacgatcct	4920
catcctgtct	cttgatcaga	tcttgatccc	ctgcgccatc	agatccttgg	cggcaagaaa	4980
gccatccagt	ttactttgca	gggcttccca	accttaccag	agggcgcccc	agctggcaat	5040
tccggttcgc	ttgctgtcca	taaaaccgcc	cagtctagct	atcgccatgt	aagcccactg	5100
caagctacct	gctttctctt	tgcgcttgcg	ttttcccttg	tccagatagc	ccagtagctg	5160
acattcatcc	ggggtcagca	ccgtttctgc	ggactggctt	tctacgtggt	ccgcttctct	5220
tagcagccct	tgcgccctga	gtgcttgccg	cagcgtgaag	ctaattcatg	gttaaatttt	5280
tgtaaataca	gctcattttt	taaccaatag	gccgaaatcg	gcaaaatccc	ttataaatca	5340
aaagaatagc	ccgagatagg	gttgagtgtt	gttccagttt	ggaacaagag	tccactatta	5400
aagaacgtgg	actccaacgt	caaagggcga	aaaaccgtct	atcagggcga	tggccggatc	5460
agcttatgcg	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccgc	atcaggcgct	5520
cttcgccttc	ctcgcctcct	gactcgtcgc	gctcggctcg	tcggctgcgg	cgagcgggat	5580
cagctcactc	aaaggcggta	atacggttat	ccacagaatc	aggggataac	gcaggaaaga	5640
acatgtgagc	aaaaggccag	caaaaggcca	ggaaccgtaa	aaaggccgcg	ttgctggcgt	5700
ttttccatag	gctccgcccc	cctgacgagc	atcacaaaaa	tcgacgctca	agtcagagg	5760
ggcgaaaccc	gacaggacta	taaagatacc	aggcgtttcc	ccctggaagc	tccctcgctg	5820
gctctcctgt	tccgaccctg	ccgcttaccg	gatacctgtc	cgcctttctc	ccttcgggaa	5880
gcgtggcgct	ttctcatagc	tcacgctgta	ggatatctcag	ttcgggtgtag	gtcgttcgct	5940
ccaagctggg	ctgtgtgcac	gaaccccccg	ttcagcccga	ccgctgcgcc	ttatccggta	6000
actatcgtct	tgagtccaac	ccggtgaagac	acgacttatc	gccactggca	gcagccactg	6060
gtaacaggat	tagcagagcg	aggtatgtag	gcggtgctac	agagttcttg	aagtgggtgg	6120
ctaactacgg	ctacactaga	aggacagtat	ttggtatctg	cgctctgctg	aagccagtta	6180
ccttcggaaa	aagagttggg	agctcttgat	ccggcaaaaa	aaccaccgct	ggtagcgggtg	6240
gtttttttgt	ttgcaagcag	cagattacgc	gcagaaaaaa	aggatctcaa	gaagatcctt	6300
tgatcttttc	tactgaacgg	tgatccccac	cggaaattgcg			6340

<210> 55
 <211> 5344
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pCMVKm2.Protopt HML-2 vector

<400>	55					
gccgcggaat	ttcgactcta	ggccattgca	tacgttgtat	ctatatcata	atatgtacat	60
ttatattggc	tcatgtccaa	tatgaccgcc	atggtgacat	tgattattga	ctagttatta	120

Substitute Sequence Listing_USSN 10587032_PP019482.007

atagtaatca	attacggggt	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggcccg	ctggctgacc	gcccacgac	ccccgccc	tgacgtcaat	240
aatgacgtat	gttcccatag	taacgccaat	agggactttc	cattgacgtc	aatgggtgga	300
gtatttacgg	taaactgccc	acttggcagt	acatcaagt	tatcatatgc	caagtccgcc	360
ccctattgac	gtcaatgacg	gtaaattggc	cgcctggcat	tatgcccagt	acatgacctt	420
acgggacttt	cctacttggc	agtacatcta	cgtattagtc	atcgctatta	ccatggtgat	480
gcggttttgg	cagtacacca	atgggcgtgg	atagcggttt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtca	atgggagttt	gttttggcac	caaaatcaac	gggactttcc	600
aaaatgtcgt	aataacccc	ccccgttgac	gcaaattggc	ggtaggcgtg	tacggtggga	660
ggtctatata	agcagagctc	gttttagtgaa	ccgtcagatc	gcctggagac	gccatccacg	720
ctgttttgac	ctccatagaa	gacaccggga	ccgatccagc	ctccgcggcc	gggaacgggtg	780
cattggaacg	cggattcccc	gtgccaagag	tgacgtaagt	accgcctata	gactctatag	840
gcacaccctt	ttggctctta	tgcatgctat	actgtttttg	gcttggggcc	tatacacc	900
cgcttcctta	tgctataggt	gatggtatag	cttagcctat	aggtgtgggt	tattgacctt	960
tattgaccac	tcccctattg	gtgacgatac	tttccattac	taatccataa	catggctctt	1020
tgccacaact	atctctattg	gctatatgcc	aatactctgt	ccttcagaga	ctgacacgga	1080
ctctgtat	ttacaggatg	gggtcccatt	tattattttac	aaattcacat	atacaacaac	1140
gccgtccccc	gtgcccgcag	tttttattaa	acatagcgtg	ggatctccac	gcgaatctcg	1200
ggtacgtgtt	ccggacatgg	gctcttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
tgggtccatg	cttccagcgg	ctcatggtcg	ctcggcagct	ccttgctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accacagatg	tgccgcacaa	ggccgtggcg	1380
gtagggtatg	tgtctgaaaa	tgagctcgga	gattgggctc	gcaccgctga	cgcagatgga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagctgag	ttgttgattt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgctgtta	acggtggagg	gcagtgtagt	ctgagcagta	1560
ctcgttgctg	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttcctt	1620
tccatgggtc	ttttctgacg	tcaccgtcgt	cgagccacc	atgtgggcca	ccatcgtggg	1680
caagcgcgcc	aagggccccc	ccagcggccc	caccaccaac	tggggcatcc	ccaacagcgc	1740
catctgcagc	agcggcttca	gcggcaccac	cacccccacc	gtgcccagcg	tgagcggcaa	1800
caagcccgtg	accaccatcc	agcagctgag	ccccgccacc	agcggcagcg	ccgccgtgga	1860
cctgtgcacc	atccaggccg	tgagcctgct	gcccggcgag	cccccccaga	agacccccac	1920
cggcgtgtac	ggccccctgc	ccaagggcac	cgtgggcctg	atcctggg	gcagcagcct	1980
gaacctgaag	ggcgtgcaga	tcacaccag	cgtggtggag	agcgactaca	agggcgagat	2040
ccagctgggtg	atcagcagca	gactccccctg	gagcggcagc	ccccgcgacc	gcatacgcca	2100
gctgctgctg	ctgccctaca	tcaagggcgg	caacagcgag	atcaagcgca	tcggcggcct	2160
gggcagcacc	gacccccacc	gcaaggccgc	ctactggg	agccagggtga	gcgagaaccg	2220
ccccgtgtgc	aaggccatca	tccagggcaa	gcagttcgag	ggcctgggtg	acaccggcgc	2280
cgacgtgagc	atcatcgccc	tgaaccagt	gcccagaagc	tgcccgaagc	agaaggccgt	2340
gaccggcctg	gtgggcatcg	gcaccgccag	cgaggtgtac	caagacaccg	agatcctgca	2400
ctgcttgggc	cccgaacaacc	aggagagcac	cgtgcagccc	atgatcacca	gcataccccct	2460
gaacctgtgg	ggccgcgacc	tgctgcagca	gtggggcgcc	gagatcacca	tgcccgcccc	2520
cagctacagc	cccaccagcc	agaagatcat	gaccaagatg	ggctacatcc	ccggcaagg	2580
cctgggcaag	aacgaggacg	gcatcaagat	ccccgtggag	gccaaagatca	accaggagcg	2640
cgagggcatc	ggcaacccct	gcgcttaag	aattcagact	cgagcaagtc	tagaaagcca	2700
tggatatcgg	atccactacg	cgttagagct	cgtcgactag	cctcgactgt	gccttctagt	2760
tgccagccat	ctgttgtttg	ccccctcccc	gtgccttcc	tgaccctgga	aggtgccact	2820
cccactgtcc	tttcttaata	aaatgaggaa	attgcatcgc	attgtctgag	taggtgtcat	2880
tctattctgg	ggggtggggt	ggggcaggac	agcaagggg	aggattggga	agacaatagc	2940
aggggggtgg	gcgaagaact	ccagcatgag	atccccgcgc	tggaggatca	tccagccggc	3000
gtcccggaaa	acgattccga	agcccaacct	ttcatagaag	gcggcgggtg	aatcgaaatc	3060
tcgtgatggc	aggttggg	tcgcttggtc	ggtcatttcg	aacccagag	tcccgtcag	3120
aagaactcgt	caagaaggcg	atagaaggcg	atgctgtcgc	aatcgggagc	ggcgataccg	3180
taaagcacga	ggaagcggtc	agcccatcgc	ccgccaagct	cttcagcaat	atcacgggta	3240
gccaacgcta	tgctctgata	gcggtccgcc	acaccagcc	ggccacagtc	gatgaatcca	3300
gaaaagcggc	cattttccac	catgatattc	ggcaagcagg	catcgccatg	ggtcacgacg	3360
agatcctcgc	cgctcgggcat	gcgcgccttg	agcctggcga	acagttcggc	tggcgcgagc	3420
ccctgatgct	cttcgtccag	atcatcctga	tcgacaagac	cggcttccat	ccgagtacgt	3480
gctcgctcga	tgcgatgttt	cgcttgggtg	tgcgaatggc	aggtagccgg	atcaagcgta	3540
tgacggcgcc	gcattgcatc	agccatgatg	gatactttct	cggcaggagc	aaggtgagat	3600
gacaggagat	cctgccccgg	cacttcgccc	aatagcagcc	agtccttcc	cgcttcagt	3660
acaacgctga	gcacagctgc	gcaagggaacg	cccgtcgtgg	ccagccacga	tagccgcgct	3720
gcctcgctcct	gcagttcatt	cagggcaccg	gacaggtcgg	tcttgacaaa	aagaaccggg	3780
cgcccctgcg	ctgacagccg	gaacacggcg	gcatacgagc	agccgattgt	ctgttgtgcc	3840
cagtcatagc	cgaatagcct	ctccacccaa	gcggcggag	aacctgcgtg	caatccatct	3900

Substitute Sequence Listing_USSN 10587032_PP019482.007

tggtcaatca	tgcgaaacga	tcctcatcct	gtctcttgat	cagatcttga	tcccctgcgc	3960
catcagatcc	ttggcggcaa	gaaagccatc	cagtttactt	tgaggggctt	cccaacctta	4020
ccagagggcg	ccccagctgg	caattccggt	tcgcttgctg	tccataaaac	cgcccagttt	4080
agctatcgcc	atgtaagccc	actgcaagct	acctgctttc	tctttgctgt	tgctgtttcc	4140
cttgtccaga	tagcccagta	gctgacattc	atccgggggtc	agcaccgttt	ctgctgactg	4200
gctttctacg	tgttccgctt	cccttagcag	cccttgctgc	ctgagtgttt	gcggcgagct	4260
gaagctaatt	catggttaaa	tttttggtta	atcagctcat	ttttaacca	ataggccgaa	4320
atcggaacaa	tcccttataa	atcaaaagaa	tagcccgaga	taggggtgag	tggtgttcca	4380
gtttggaaca	agagtccact	attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaacc	4440
gtctatcagg	gcgatggccg	gatcagctta	tgcggtgtga	aataccgcac	agatgcgtaa	4500
ggagaaaata	ccgcatcagg	cgctcttccg	cttcctcgct	cactgactcg	ctgctgtcgg	4560
tcgttcggct	gcggcgagcg	gtatcagctc	actcaaaggc	ggtaatacgg	ttatccacag	4620
aatcagggga	taacgcagga	agaacatgt	gagcaaaagg	ccagcaaaag	gccaggaacc	4680
gtaaaaaggc	cgctgtgctg	gcgtttttcc	ataggctccg	ccccctgac	gagcatcaca	4740
aaaaatcgac	ctcaagtcag	aggtggcgaa	acccgacagg	actataaaga	taccaggcgt	4800
ttccccctgg	aagctccctc	gtgcgtcttc	ctgttccgac	cctgccgctt	accggatacc	4860
tgctccgctt	tctcccttcg	ggaagcgtgg	cgctttctca	tagctcacgc	tgtaggtatc	4920
tcagttcggt	gtaggtcggt	cgctccaagc	tgggctgtgt	gcacgaaccc	cccgttcagc	4980
ccgaccgctg	cgcttatcc	ggtaactatc	gtcttgagtc	caaccgggta	agacacgact	5040
tatcgccact	ggcagcagcc	actggtaaca	ggatttagac	agcgaggtat	gtaggcggtg	5100
ctacagagtt	cttgaagtgg	tggcctaact	acggctacac	tagaaggaca	gtatttggtg	5160
tctgcgctct	gctgaagcca	gttaccttcg	gaaaaagagt	tggtagctct	tgatccggca	5220
aacaaaccac	cgctggtagc	ggtggttttt	ttgtttgcaa	gcagcagatt	acgcgcagaa	5280
aaaaaggatc	tcaagaagat	ctttgatct	tttctactga	acggtgatcc	ccaccggaat	5340
tgcg						5344

<210> 56
 <211> 7211
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pCMVKm2.Polopt HML-2 vector

<400>	56					
gccgcggaat	ttcgactcta	ggccattgca	tacgttgtat	ctatatcata	atatgtacat	60
ttatattggc	tcatgtccaa	tatgaccgcc	atgttgacat	tgattattga	ctagttatta	120
atagtaatca	attacggggt	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggcccgc	ctggctgacc	gcccacagac	ccccgccc	tgacgtcaat	240
aatgacgtat	gttcccatag	taacgccaat	agggactttc	cattgacgtc	aatgggtgga	300
gtattttacg	taaactgccc	acttggcagt	acatcaagt	tatcatatgc	caagtccgcc	360
ccctattgac	gtcaatgacg	gtaaatggcc	cgcttgcat	tatgccaggt	acatgacctt	420
acgggacttt	cctacttggc	agtacatcta	cgtattagtc	atcgctatta	ccatggtgat	480
gcggttttgg	cagtacacca	atgggcgtgg	atagcggttt	gactcacggg	gattttccaag	540
tctccacccc	attgacgtca	atgggagttt	gttttggcac	caaaatcaac	gggactttcc	600
aaaatgtcgt	aataaccccc	ccccgttgac	gcaaatgggc	ggtaggcgtg	tacggtggga	660
ggtctatata	agcagagctc	gtttagtga	ccgtcagatc	gcctggagac	gccatccacg	720
ctgttttgac	ctccatagaa	gacaccggga	ccgatccagc	ctccgcggcc	gggaacggtg	780
cattggaacg	cggattcccc	gtgccaaag	tgacgtaagt	accgcctata	gactctatag	840
gcacacccct	ttggctctta	tgcatgctat	actgtttttg	gcttggggcc	tatacacccc	900
cgcttcctta	tgctataggt	gatggtatag	cttagcctat	aggtgtgggt	tattgacctt	960
tattgaccat	tcccctattg	gtgacgatac	tttccattac	taatccataa	catggctctt	1020
tgccacaact	atctctattg	gctatatgcc	aatactctgt	ccttcagaga	ctgacacgga	1080
ctctgtattt	ttacaggatg	gggtcccatt	tattattttac	aaattcacat	atacaacaac	1140
gccgtccccc	gtgcccgag	tttttattaa	acatagcgtg	ggatctccac	gcgaatctcg	1200
ggtacgtgtt	ccggacatgg	gctcttctcc	ggttagcggc	gagcttccac	atccgagccc	1260
tgggtcccat	ctccagcgg	ctcatggtcg	ctcggcagct	ccttgctcct	aacagtggag	1320
gccagactta	ggcacgcac	aatgccacc	accaccagtg	tgccgcacaa	ggccgtggcg	1380
gtagggtatg	tgtctgaaaa	tgagctcgga	gattgggctc	gcaccgctga	cgcagatgga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagctgag	ttgttgattt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgctgtta	acggtggagg	gcagtgtagt	ctgagcagta	1560
ctcgttgctg	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttcctt	1620
tccatgggtc	ttttctgcag	tcaccgtcgt	cgacgccacc	atgaacaaga	gccgcaagcg	1680

Substitute Sequence Listing_USSN 10587032_PP019482.007

ccgcaaccgc	gagagcctgc	tgggcgccgc	caccgtggag	cccccaagc	ccatccccct	1740
gacctggaag	accgagaagc	ccgtgtgggt	gaaccagtgg	cccctgcca	agcagaagct	1800
ggaggccctg	cacctgtctg	ccaacgagca	gctggagaag	ggccacatcg	agcccagctt	1860
cagccccctg	aacagccccg	tgttcgtgat	ccagaagaag	agcggcaagt	ggcgcatgct	1920
gaccgacctg	cgcgccgtga	acgccgtgat	ccagcccatg	ggccccctgc	agcccggcct	1980
gcccagcccc	gccatgatcc	ccaaggactg	gccccctgatc	atcatcgacc	tgaaggactg	2040
cttcttcacc	atccccctgg	ccgagcagga	ctgcgagaag	ttcgccttca	ccatccccgc	2100
catcaacaac	aaggagcccg	ccacccgctt	ccagtggaa	gtgctgcccc	agggcatgct	2160
gaacagcccc	accatctgcc	agaccttcgt	gggcccgcgc	ctgcagcccc	tgcgcgagaa	2220
gttcagcgac	tgctacatca	tccactgcat	cgacgacatc	ctgtgcgccg	ccgagaccaa	2280
ggacaagctg	atcgactgct	acaccttcct	gcaggccgag	gtggccaacg	ccggcctggc	2340
catcgccagc	gacaagatcc	agaccagcac	ccccttccac	tacctgggca	tgcagatcga	2400
gaaccgcaag	atcaagcccc	agaagatcga	gatccgcaag	gacaccctga	agaccctgaa	2460
cgacttccag	aagctgtctg	gcgacatcaa	ctggatccgc	cccacccctg	gcatccccac	2520
ctacgccatg	agcaacctgt	tcagcatcct	gcgcggcgac	agcgacctga	acagcaagcg	2580
catgctgacc	cccagggccca	ccaaggagat	caagctggtg	gaggagaaga	tccagagcgc	2640
ccagatcaac	cgcatcgacc	ccctggcccc	cctgcagctg	ctgatcttcg	ccaccgcccc	2700
cagccccacc	ggcatcatca	tccagaacac	cgacctggtg	gagtggagct	tcctgccccca	2760
cagcaccgtg	aagaccttca	ccctgtacct	ggaccagatc	gccaccctga	tcggccagac	2820
ccgcctgcgc	atcatcaagc	tgtgcggcaa	cgaccccgcg	aagatcgtgg	tgccccctgac	2880
caaggagcag	gtgcgcagcg	cttctatcaa	cagcggcgcc	tggaagatcg	gcctggccaa	2940
cttcgtgggc	atcatcgaca	accactaccc	caagaccaag	atcttccagt	tcctgaagct	3000
gaccacctgg	atcctgcccc	agatcacccg	ccgcgagccc	ctggagaacg	ccctgaccgt	3060
gttcaccgac	ggcagcagca	acggcaaggc	cgcctacacc	ggcccccaagg	agcgcgtgat	3120
caagaccccc	taccagagcg	cccagcgcgc	cgagctggtg	gccgtgatca	ccgtgtctga	3180
ggacttcgag	cagcccatca	acatcatcag	cgacagcgcc	tacgtggtgc	aggccaccgc	3240
cgagctggag	accgcccgtg	tcaagtacag	catggacgac	cagctgaacc	agctgttcaa	3300
cctgctgcag	cagaccgtgc	gcaagcgcaa	cttccccctc	tacatcaccc	acatccgcgc	3360
ccacaccaac	ctgcccggcc	ccctgaccaa	ggccaacgag	caggccgacc	tgctggtgag	3420
cagcgccctg	atcaaggccc	aggagctgca	cgccctgacc	cacgtgaacg	ccgcccgcct	3480
gaagaacaag	ttcgacgtga	cctggaagca	ggccaaggac	atcgtgcagc	actgcaccca	3540
gtgccagggtg	ctgcacctgc	ccacccagga	ggcggcgctg	aacccccgcg	gcctgtgccc	3600
caacgcccctg	tggcagatgg	acgtgaccca	ctgtgcccgc	ttcggccgcc	tgagctacgt	3660
gcacgtgacc	gtggacacct	acagccactt	catctggggc	acctgccaga	ccggcgagag	3720
caccagccac	gtgaagaagc	acctgctgag	ctgcttcgcc	gtgatgggcg	tgcccgagaa	3780
gatcaagacc	gacaacggcc	ccggctactg	cagcaaggcc	ttccagaagt	tcctgagcca	3840
gtggaagatc	agccacacca	ccggcatccc	ctacaacagc	cagggccagg	ccatcgtgga	3900
gcgaccaaac	cgcaccctga	agacccagct	ggtgaagcag	aaggaggggcg	gcgacagcaa	3960
ggagtgcacc	accccccaga	tgacgtgaa	cctggcccctg	tacaccctga	acttcctgaa	4020
catctaccgc	aaccagacca	ccaccagcgc	cgacgagcac	ctgaccggca	agaagaacag	4080
cccccacgag	ggcaagctga	tctggtggaa	ggacaacaag	aacaagacct	gggagatcgg	4140
caaggtgatc	acctggggcc	gcggcttcgc	ctgctgtgagc	cccggcgaga	accagctgcc	4200
cgtgtggatc	cccacccgcc	acctgaagtt	ctacaacgag	cccattccgcg	acgccaagaa	4260
gagcaccagc	gccgagaccg	agaccagcca	gagcagcacc	gtggacagcc	aggacgagca	4320
gaacggcgac	gtgcgcccga	ccgacgaggt	ggccatccac	caggaggggc	gcgcccga	4380
cctggggcacc	accaaggagg	ccgacgcggt	gagctacaag	atcagccgcg	agcacaaggg	4440
cgacaccaac	ccccgcgagt	acgcccgcctg	cagcctggac	gactgcatca	acggcgga	4500
gagccccctac	gcctgccgca	gcagctgcag	cttaagaagt	tcagactcga	gcaagtctag	4560
aaagccatgg	atatcggtatc	cactacgcgt	tagagctcgc	tgatcagcct	cgactgtgcc	4620
ttctagtgtg	cagccatctg	ttgtttgccc	ctccccctg	ccttccttga	ccctggaagg	4680
tgccactccc	actgtccttt	cctaataaaa	tgaggaaatt	gcacccgatt	gtctgagtag	4740
gtgtcattct	attctggggg	gtgggggtgg	gcaggacagc	aagggggagg	attgggaaga	4800
caatagcagg	ggggtgggcg	aagaactcca	gcatgagatc	cccgcgctgg	aggatcatcc	4860
agccggcgtc	ccggaaaacg	attccgaagc	ccaacctttc	atagaaggcg	gcggtggaat	4920
cgaaatctcg	tgatggcagg	ttgggcgctg	cttggtcggt	catttcgaac	cccagagtcc	4980
cgctcagaag	aactcgtcaa	gaaggcgata	gaaggcgatg	cgctgcgaat	cgggagcggc	5040
gataccgtaa	agcacgagga	agcggtcagc	ccattccgcc	ccaagctctt	cagcaatatc	5100
acgggtagcc	aacgctagt	cctgatagcg	cttcgcccga	cccagccggc	cacagtcgat	5160
gaatccagaa	aagcggccat	tttccaccat	gatattcggc	aagcaggcat	cgccatgggt	5220
cacgacgaga	tcctcgccgt	cgggcatgcg	cgccttgagc	ctggcgaaca	gttcggctgg	5280
cgcgagcccc	tgatgctctt	cgtccagatc	atcctgatcg	acaagaccgg	cttccatccg	5340
agtacgtgct	cgctcgatgc	gatgtttcgc	ttggtggtcg	aatgggcagg	tagccggatc	5400
aagcgtatgc	agccgcccga	ttgcatcagc	catgatggat	actttctcgg	caggagcaag	5460

Substitute Sequence Listing_USSN 10587032_PP019482.007

gtgagatgac	aggagatcct	gccccggcac	ttcgcccaat	agcagccagt	cccttcccgc	5520
ttcagtgaca	acgtcagagca	cagctgcgca	aggaacgccc	gtcgtggcca	gccacgatag	5580
ccgcgctgcc	tcgtcctgca	gttcattcag	ggcaccggac	aggctcgtct	tgacaaaaag	5640
aaccggggcg	ccctgcgctg	acagccggaa	cacggcggca	tcagagcagc	cgattgtctg	5700
ttgtgcccag	tcatagccga	atagcctctc	cacccaagcg	gccggagaac	ctgctgtcaa	5760
tccatcttgt	tcaatcatgc	gaaacgatcc	tcacctctgt	tcttgatcag	atcttgatcc	5820
cctgcgccat	cagatccttg	gcggcaagaa	agccatccag	tttactttgc	agggcttccc	5880
aaccttacca	gagggcgccc	cagctggcaa	ttcggttcgc	cttgctgtcc	ataaaaaccgc	5940
ccagtctagc	tatcgccatg	taagcccact	gcaagctacc	tgctttctct	ttgcgcttgc	6000
gttttccctt	gtccagatag	cccagtagct	gacattcatc	cggggtcagc	accgtttctg	6060
cggactggct	ttctacgtgt	tccgcttcc	ttagcagccc	ttgcgccctg	agtgttgcg	6120
gcagcgtgaa	gctaattcat	ggttaaattt	ttgttaaatt	agctcatttt	ttaaccaata	6180
ggccgaaatc	ggcaaaatcc	cttataaatc	aaaagaatag	cccagatag	ggttgagtgt	6240
tgttccagtt	tgaacaaga	gtccactatt	aaagaacgtg	gactccaacg	tcaaaggcg	6300
aaaaaccgtc	tatcaggcg	atggccggat	cagcttatgc	gggtgtgaaat	accgcacaga	6360
tgcgtaagga	gaaaataccg	catcaggcgc	tcttccgctt	cctcgctcac	tgactcgtg	6420
cgctcggctg	ttcggctgcg	gcgagcggta	tcagctcact	caaaggcgg	aatacgggta	6480
tccacagaa	caggggataa	cgcaggaaag	aacatgtgag	caaaaggcca	gcaaaaggcc	6540
aggaaccgta	aaaaggccgc	gttgctggcg	tttttccata	ggctccgccc	ccctgacgag	6600
catcacaaaa	atcgacgctc	aagtcagagg	tggcgaaacc	cgacaggact	ataaagatac	6660
caggcggttt	cccctggaa	ctccctcgtg	cgctctcctg	ttccgaccct	gccgcttacc	6720
ggatacctgt	ccgcctttct	cccttcggga	agcgtggcgc	tttctcatag	ctcacgctgt	6780
aggtatctca	gttcggtgta	ggcgttcgc	tccaagctgg	gctgtgtgca	cgaaccccc	6840
gttcagccc	accgctgcgc	cttatccggt	aactatcgtc	ttgagtccaa	cccggtaaga	6900
cacgacttat	cgccactggc	agcagccact	ggtaacagga	ttagcagagc	gaggtatgta	6960
ggcgggtgta	cagagttctt	gaagtgttgg	cctaactacg	gctacactag	aaggacagta	7020
tttggtatct	gcgctctgct	gaagccagtt	accttcggaa	aaagagttgg	tagctcttga	7080
tccggcaaac	aaaccaccgc	tggtagcgg	ggttttttt	tttgcaagca	gcagattacg	7140
cgcagaaaa	aaggatctca	agaagatcct	ttgatctttt	ctactgaacg	gtgatcccca	7200
ccggaattgc	g					7211

<210> 57

<211> 318

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 57

atgaacccat	cagagatgca	aagaaaagca	cctccgcgga	gacggagaca	tcgcaatcga	60
gcaccggttg	ctcacaaagat	gaacaaaatg	gtgacgtcag	aagaacagat	gaagttgcca	120
tccaccaaga	aggcagagcc	gccaaacttg	gcacaactaa	agaagctgac	gcagttagct	180
acaaaatatc	tagagaacac	aaaggtgaca	caaaccctag	agagtatgct	gcttgacgac	240
ttgatgattg	tatcaatgg	gtctgcaggt	gtaccaaca	gctccgaaga	gacagcgacc	300
atcgagaacg	ggccatga					318

<210> 58

<211> 321

<212> DNA

<213> Artificial Sequence

<220>

<223> Modified cORF sequence

<400> 58

atgaacccca	gcgagatgca	gcgcaaggcc	cccccccgcc	gccgccgcca	ccgcaaccgc	60
gccccctga	cccacaagat	gaacaagatg	gtgaccagcg	aggagcagat	gaagctgccc	120
agcaccaaga	aggccgagcc	ccccacctgg	gcccagctga	agaagctgac	ccagctggcc	180
accaagtacc	tggagaacac	caaggtgacc	cagacccccg	agagcatgct	gctggccgcc	240
ctgatgatcg	tgagcatggt	gagcgcggcg	gtgcccaaca	gcagcagagga	gaccgcccacc	300
atcgagaacg	gccccgctta	a				321

<210> 59

<211> 435

<212> DNA

Substitute Sequence Listing_USSN 10587032_PP019482.007

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 59

atgaacccat	cggagatgca	aagaaaagca	cctccgcgga	gacggagaca	tcgcaatcga	60
gcaccgttga	ctcacaagat	gaacaaaatg	gtgacgtcag	aagaacagat	gaagttgcca	120
tccaccaaga	aggcagagcc	gccaaacttg	gcacaactaa	agaagctgac	gcagtttagct	180
acaaaatc	tagagaacac	aaaggtgaca	caaaccaccg	agagtatgct	gcttgacagcc	240
ttgatgattg	tatcaatggg	gggtgtaccca	acagctccga	agagacagcg	accatcgaga	300
acgggccatg	atgacgatgg	cggttttgtc	gaaaagaaaa	gggggaaatg	tggggaaaaag	360
caagagagat	cagattgtta	ctgtgtctgt	gtagaaagaa	gtagacatag	gagactccat	420
ttgttctgt	actaa					435

<210> 60

<211> 438

<212> DNA

<213> Artificial Sequence

<220>

<223> Modified PCAP5 sequence

<400> 60

atgaacccca	gcgagatgca	gcgcaaggcc	cccccccgcc	gccgccgcca	ccgcaaccgc	60
gccccctga	cccacaagat	gaacaagatg	gtgaccagcg	aggagcagat	gaagctgccc	120
agcaccaaga	aggccgagcc	ccccacctgg	gcccagctga	agaagctgac	ccagctggcc	180
accaagtacc	tggagaacac	caaggtgacc	cagacccccg	agagcatgct	gctggccgcc	240
ctgatgatcg	tgagcatggg	gggtgtaccc	accgccccca	agcggcagcg	ccccagccgc	300
accggccatg	acgacgagcg	cggttctgtg	gagaagaagc	gcggcaagtg	cggcgagaag	360
caggagcgca	gcgactgcta	ctgcgtgtgc	gtggagcgca	gccgccaccg	ccgcctgcac	420
ttcgtgctgt	acgcttaa					438

<210> 61

<211> 2001

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 61

atggggcaaa	ctaaaagtaa	aattaaaagt	aaatatgcct	cttatctcag	ctttattaaa	60
attcttttaa	aaagaggggg	agttaaagta	tctacaaaaa	atctaataca	gctattttcaa	120
ataatagaac	aattttgccc	atggttttcca	gaacaaggaa	ctttagatct	aaaagatttg	180
aaaagaattg	gtaagggaact	aaaacaagca	ggtaggaagg	gtaatatcat	tccactttaca	240
gtatggaatg	attgggccat	tattaaagca	gctttagaac	cattttcaaac	agaagaagat	300
agcgtttcag	tttctgatgc	ccctggaagc	tgtataatag	attgtaatga	aaacacaagg	360
aaaaaatccc	agaaaagaaac	ggaagggtta	cattgccaat	atgtagcaga	gccggtaaatg	420
gctcagtaaa	cgcaaaatgt	tgactataat	caattacagg	aggtgatata	tcctgaaacg	480
ttaaaattag	aaggaaaagg	tccagaatta	gtggggccat	cagagtctaa	accacgaggc	540
acaagtccct	ttccagcagg	tcagggtgct	gtaacattac	aacctcaaaa	gcagggttaa	600
gaaaataaga	cccaaccgcc	agtagcctat	caatactggc	ctccggctga	acttcagtat	660
cggccacccc	cagaaaagtca	gtatggatat	ccaggaatgc	ccccagcacc	acagggcagg	720
gcgccatacc	ctcagccgcc	cactaggaga	cttaatccta	cggcaccacc	tagtagacag	780
ggtagtaaat	tacatgaaat	tattgataaa	tcaagaaagg	aaggagatac	tgaggcatgg	840
caattcccag	taacgttaga	accgatgcc	cctggagaag	gagcccaaga	gggagagcct	900
cccacagtgg	aggccagata	caagtctttt	tcgataaaaa	agctaaaaga	tatgaaagag	960
ggagtataaac	agtatggacc	caactcccc	tatatgagga	cattattaga	ttccattgct	1020
catggacata	gactcattcc	ttatgattgg	gagattctgg	caaaatcgct	tctctcacc	1080
tctcaatttt	tacaatttaa	gacttggtgg	attgatgggg	tacaagaaca	gggtccgaaga	1140
aatagggtcg	ccaatcctcc	agttaacata	gatgcagatc	aactattagg	aatagggtcaa	1200
aattggagta	ctattagtca	acaagcata	atgcaaaatg	aggccattga	gcaagttaga	1260
gctatctgcc	ttagagccctg	ggaaaaaatc	caagaccagg	gaagtacctg	ccctcatttt	1320
aatacagtaa	gacaagggtc	aaaagagccc	tatcctgatt	ttgtggcaag	gctccaagat	1380
gttgctcaaa	agtcaattgc	tgatgaaaaa	gcccgttaag	tcatagtgga	gttgatggca	1440
tatgaaaacg	ccaatcctga	gtgtcaatca	gccattaagc	cattaaaagg	aaagggttct	1500
gcaggatcag	atgtaatctc	agaatatgta	aaagcctgtg	atggaatcgg	aggagctatg	1560
cataaagcta	tgcttatggc	tcaagcaata	acaggagtgg	ttttaggagg	acaagttaga	1620

Substitute Sequence Listing_USSN 10587032_PP019482.007

acatttggaa	gaaaatgtta	taattgtggt	caaattgggtc	acttaaaaaa	gaattgcca	1680
gtcttaaata	aacagaatat	aactattcaa	gcaactacaa	caggtagaga	gccacctgac	1740
ttatgtccaa	gatgtaaaaa	aggaaaacat	tgggctagtc	aatgtcgttc	taaatttgat	1800
aaaaatgggc	aaccattgtc	gggaaacgag	caaagggggc	agcctcaggc	cccacaacaa	1860
actggggcat	tcccaattca	gccatttgtt	cctcagggtt	ttcaggggaca	acaacccccca	1920
ctgtcccaag	tgtttcaggg	aataagccag	ttaccacaat	acaacaattg	tcccccgcca	1980
caagcggcag	tgacagcagta	g				2001

<210> 62
 <211> 2004
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified gag sequence

<400> 62						
atggggccaga	ccaagagcaa	gatcaagagc	aagtacgcca	gctacctgag	cttcatcaag	60
atcctgctga	agcgcggcgg	cgtgaagggtg	agcaccaaga	acctgatcaa	gctgttccag	120
atcatcgagc	agttctgccc	ctggttcccc	gagcagggca	ccctggacct	gaaggactgg	180
aagcgcatcg	gcaaggagct	gaagcaggcc	ggccgcaagg	gcaacatcat	ccccctgacc	240
gtgtggaacg	actgggccat	catcaaggcc	ggcctggagc	ccttccagac	cgaggaggac	300
agcgtgagcg	tgagcgacgc	ccccggcagc	tgcatcatcg	actgcaacga	gaaacaccgc	360
aagaagagcc	agaaggagac	cgagggcctg	cactgcgagt	acgtggccga	gcccgtgatg	420
gcccagagca	cccagaacgt	ggactacaac	cagctgcagg	aggtgatcta	ccccgagacc	480
ctgaagctgg	agggcaaggg	ccccgagctg	gtggggccca	gcgagagcaa	gccccgcggc	540
accagccccc	tgcccgcggg	ccaggtgccc	gtgaccctgc	agccccagaa	gcaggtgaag	600
gagaacaaga	cccagccccc	cgtggcctac	cagtactggc	ccccgcgca	gctgcagtac	660
cgcccccccc	ccgagagcca	gtacggctac	cccggcatgc	ccccgcccc	ccagggccgc	720
gccccctacc	cccagccccc	caccgcggc	ctgaacccca	ccgccccccc	cagccgccag	780
ggcagcaagc	tgacagagat	catcgacaag	agccgcaagg	agggcgacac	cgaggcctgg	840
cagttccccg	tgaccctgga	gccccatgcc	cccggcgagg	gcgcccagga	gggcgagccc	900
cccaccgtgg	aggcccgccta	caagagcttc	agcatcaaga	agctgaagga	catgaaggag	960
ggcgtgaagc	agtacggccc	caacagcccc	tacatgcgca	ccctgctgga	cagcatcgcc	1020
cacggccacc	gcctgatccc	ctacgactgg	gagatcctgg	ccaagagcag	cctgagcccc	1080
agccagttcc	tgacagttcaa	gacctgggtg	atcgacggcg	tgacaggagca	ggtgcgccgc	1140
aaccgcgccc	ccaaccccc	cgtgaacatc	gacgccgacc	agctgctggg	catcggccag	1200
aactggagca	ccatcagcca	gcaggccctg	atgcagaacg	aggccatcga	gcaggtgcgc	1260
gccatctgcc	tgcgcgccctg	ggagaagatc	caggaccccc	gcagcacctg	ccccagcttc	1320
aacaccgtgc	gccaggggcag	caaggagccc	taccccgact	tcgtggcccc	cctgcaggac	1380
gtggcccaaga	agagcatcgc	cgacgagaag	gcccgcgaag	tgatcgtgga	gctgatggcc	1440
tacgagaacg	ccaacccccga	gtgccagagc	gccatcaagc	ccctgaaggg	caaggtgccc	1500
gccggcagcg	acgtgatcag	cgagtacgtg	aaggcctgcg	acggcatcgg	cggcgccatg	1560
cacaaggcca	tgctgatggc	ccaggccatc	accggcgtgg	tgctggggcg	ccaggtgcgc	1620
accttcggcc	gcaagtgtcta	caactgcggc	cagatcggcc	acctgaagaa	gaactgcccc	1680
gtgctgaaca	agcagaacat	caccatccag	gccaccacca	ccggcccgca	gccccccgac	1740
ctgtgcccc	gctgcaagaa	gggcaagcac	tgggccagcc	agtgccgcag	caagtccgac	1800
aagaacggcc	agccccctgag	cggaacagag	cagcgcggcc	agccccaggc	ccccagcag	1860
accggcgcc	tccccatcca	gccccttcgtg	ccccagggtc	tccaggggca	gcagcccccc	1920
ctgagccagg	tggtccaggg	catcagccag	ctgccccagt	acaacaactg	cccccccccc	1980
caggccgccc	tgacagcaggc	ttaa				2004

<210> 63
 <211> 1005
 <212> DNA
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 63						
atgtgggcaa	ccattgtcgg	gaaacgagca	aagggggccag	cctcaggccc	cacaacaaac	60
tggggcattc	ccaattcagc	catttgttcc	tcagggtttt	cagggaacaac	aacccccact	120
gtcccaagtg	tttcaggga	taagccagtt	accacaatac	aacaattgtc	ccccgccaca	180
agcggcagtg	cagcagtaga	tttatgtact	atacaagcag	tctctctgct	tccaggggag	240
ccccacaaa	aaacccccac	aggggtatat	ggacccctgc	ctaaggggac	tgtaggacta	300

Substitute Sequence Listing_USSN 10587032_PP019482.007

atcttgggac	gatcaagtct	aaatctaaaa	ggagttcaaa	ttcatactag	tgtggttgat	360
tcagactata	aaggcgaaat	tcaattgggt	attagctctt	caattccttg	gagtgccagt	420
ccaagagaca	ggattgctca	attattactc	ctgccataca	ttaaggggtg	aaatagtga	480
ataaaaagaa	taggagggct	tggaagcact	gatccaacag	gaaaggctgc	atattgggca	540
agtcagggtct	cagagaacag	acctgtgtgt	aaggccatta	ttcaaggaaa	acagtttgaa	600
gggttggttag	acactggagc	agatgtctct	atcattgtct	taaatcagtg	gccaaaaaat	660
tggcctaaac	aaaaggctgt	tacaggactt	gtcggcatag	gcacagcctc	agaagtgtat	720
caaagtacgg	agattttaca	ttgcttaggg	ccagataatc	aagaaagtac	tgttcagcca	780
atgattactt	caattcctct	taatctgtgg	ggctcgagatt	tattacaaca	atgggggtgcg	840
gaaatcacca	tgcccgtctc	atcatatagc	cccacgagtc	aaaaaatcat	gaccaagatg	900
ggatatatac	caggaaaggg	actagggaaa	aatgaagatg	gcattaaaat	tccagttgag	960
gctaaaataa	atcaagaaag	agaaggaata	gggaatcctt	gctag		1005

<210> 64
 <211> 1008
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified Prt sequence

<400> 64						
atgtgggcca	ccatcgtggg	caagcgcgcc	aagggccccg	ccagcggccc	caccaccaac	60
tggggcatcc	ccaacagcgc	catctgcagc	agcggcttca	gcggcaccac	cacccccacc	120
gtgcccagcg	tgagcggcaa	caagcccgtg	accaccatcc	agcagctgag	ccccgccacc	180
agcggcagcg	ccgcccgtga	cctgtgcacc	atccaggccg	tgagcctgct	gccccgagcg	240
cccccccaga	agacccccac	cggcgtgtac	ggccccctgc	ccaagggcac	cgtgggcctg	300
atcctggggc	gcagcagcct	gaacctgaag	ggcgtgcaga	tccacaccag	cgtgggtggac	360
agcgactaca	agggcgagat	ccagctgggtg	atcagcagca	gcatccccctg	gagcgccagc	420
ccccgcgacc	gcatcgccca	gctgctgctg	ctgccctaca	tcaagggcgg	caacagcgag	480
atcaagcgca	tcggcggcct	gggcagcacc	gacccccacc	gcaaggccgc	ctactggggc	540
agccaggtga	gcgagaaccg	ccccgtgtgc	aaggccatca	tccagggcaa	gcagttcgag	600
ggcctgggtg	acaccggcgc	cgacgtgagc	atcatcgccc	tgaaccagtg	gcccagaagac	660
tggcccaagc	agaaggccgt	gaccggcctg	gtgggcatcg	gcaccgccag	cgaggtgtac	720
cagagcaccg	agatcctgca	ctgcctgggc	cccgacaacc	aggagagcac	cgtgcagccc	780
atgatcacca	gcatccccct	gaacctgtgg	ggccgcgacc	tgctgcagca	gtggggcgcc	840
gagatcacca	tgcccgcgcc	cagctacagc	cccaccagcc	agaagatcat	gaccaagatg	900
ggctacatcc	ccggcaaggg	cctgggcaag	aacgaggacg	gcatcaagat	ccccgtggag	960
gccaagatca	accaggagcg	cgagggcatc	ggcaaccctt	gcgcttaa		1008

<210> 65
 <211> 2874
 <212> DNA
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 65						
atgaataaat	caagaaagag	aaggaatagg	gaatccttgc	taggggaggc	caactgtagag	60
cctcctaaac	ccataccatt	aacttggaag	acagaaaaac	cagtgtgggt	aaatcagtg	120
ccgctacca	aacaaaaact	ggaggcttta	catttattag	caaataaaca	gttagaaaa	180
ggtcatattg	agccttcgtt	ctcaccttgg	aattctcctg	tgtttgtaat	tcagaagaaa	240
tcaggcaaat	ggcgtatgtt	aactgactta	agggctgtaa	acgccgtaat	tcaacccatg	300
gggcctctcc	aacccgggtt	gccctctccg	gccatgatcc	caaaagattg	gcctttaatt	360
ataattgatc	taaaggattg	cttttttacc	atccctcttg	cagagcagga	ttgcgaaaaa	420
tttgctttta	ctataccagc	cataaataat	aaagaaccag	ccaccaggtt	tcagtggaaa	480
gtgttacctc	agggaaatgct	taatagtcca	actattttgt	agacttttgt	aggctcgagct	540
cttcaaccag	ttagagaaaa	gttttcagac	tggttatatta	ttcattgtat	tgatgatatt	600
ttatgtgctg	cagaaacgaa	agataaatta	attagctgtt	atacatttct	gcaagcagag	660
gttgccaatg	ctggactggc	aatagcatct	gataagatcc	aaacctctac	tccttttcat	720
tatttaggga	tgcagataga	aaatagaaaa	attaagccac	aaaaaataga	aataagaaaa	780
gacacattaa	aaacactaaa	tgattttcaa	aaattactag	gagatattaa	ttggattcgg	840
ccaactctag	gcatttcctac	ttatgccatg	tcaaatttgt	tctctatctt	aagaggagac	900
tcagacttaa	atagtaaaa	aatgttaacc	ccagaggcaa	caaaagaaat	taaattagtg	960
gaagaaaaaa	ttcagtcagc	gcaaataaat	agaatagatc	ccttagcccc	actccaactt	1020

Substitute Sequence Listing_USSN 10587032_PP019482.007

ttgatttttg	ccactgcaca	ttctccaaca	ggcatcatta	ttcaaaatac	tgatcttgtg	1080
gagtgggtcat	tccttcctca	cagtacagtt	aagactttta	cattgtactt	ggatcaaata	1140
gctacattaa	tcggtcagac	aagattacga	ataataaaat	tatgtgggaa	tgacccagac	1200
aaaatagttg	tccctttaac	caaggaacaa	gttagacaag	cctttatcaa	ttctggtgca	1260
tggaagattg	gtcttgctaa	ttttgtggga	attattgata	atcattaccc	aaaaacaaag	1320
atcttccagt	tcttaaaatt	gactacttgg	attctacctt	aaattaccag	acgtgaacct	1380
ttagaaaaatg	ctctaacagt	atctactgat	ggttccagca	atggaaaagc	agcttacaca	1440
ggaccgaaaag	aacgagtaat	caaaactcca	tatcaatcgg	ctcaaagagc	agagttgggt	1500
gcagtcatta	cagtgttaca	agattttgac	caacctatca	atattatatc	agattctgca	1560
tatgtagtac	aggctacaag	ggatgttgag	acagctctaa	ttaaataatag	catggatgat	1620
cagttaaacc	agctattcaa	tttattacaa	caaactgtaa	gaaaaagaaa	tttcccattt	1680
tatattacac	atattcgagc	acacactaat	ttaccagggc	ctttgactaa	agcaaatgaa	1740
caagctgact	tactgggtac	atctgcactc	ataaaagcac	aagaacttca	tgctttgact	1800
catgtaaatg	atcagagatt	aaaaaacaaa	tttgatgtca	catggaaaca	ggcaaaagat	1860
attgtacaac	attgcaccca	gtgtcaagtc	ttacacctgc	ccactcaaga	ggcaggagtt	1920
aatcccagag	gtctgtgtcc	taatgcatta	tggcaaatgg	atgtcacgca	tgtaccttca	1980
tttggaagat	tatcatatgt	tcacgtaaca	gttgatactt	attcacattt	catatgggca	2040
acttgccaaa	caggagaaaag	tacttcccat	gttaaaaaac	atctattgtc	ttgttttgct	2100
gtaatgggag	ttccagaaaa	aatcaaaaact	gacaatggac	caggatattg	tagtaaagct	2160
ttccaaaaat	tcttaagtca	gtggaataat	tcacatacaa	caggaattcc	ttataattcc	2220
caaggacagg	ccatagttga	aagaactaat	agaacactca	aaactcaatt	agttaaacaa	2280
aaagaagggg	gagacagtaa	ggagtgtacc	actcctcaga	tgcaacttaa	tctagcactc	2340
tatactttaa	atctttttaa	catttataga	aatcagacta	ctacttctgc	agaacaacat	2400
cttactggta	aaaagaacag	cccacatgaa	ggaaaactaa	tttggtggaa	agataataaa	2460
aataagacat	gggaaatagg	gaaggtgata	acgtggggga	gaggttttgc	ttgtgtttca	2520
ccaggagaaa	atcagcttcc	tgtttgata	cccactagac	atttgaagtt	ctacaatgaa	2580
cccacatgag	atgcaaaagaa	aagcacctcc	gcggagacgg	agacatcgca	atcgagcacc	2640
gttgactcac	aagatgaaca	aaatggtgac	gtcagaagaa	cagatgaagt	tgccatccac	2700
caagaaggca	gagccgcaa	cttgggcaca	actaaagaag	ctgacgcagt	tagctacaaa	2760
atatctagag	aacacaaagg	tgacacaaac	cccagagagt	atgctgcttg	cagccttgat	2820
gattgtatca	atggtggtaa	gtctccctat	gcctgcagga	gcagctgcag	ctaa	2874

<210> 66
 <211> 2877
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified pol sequence

<400>	66					
atgaacaaga	gccgcaagcg	ccgcaaccgc	gagagcctgc	tgggcgcgcg	caccgtggag	60
cccccaagc	ccatccccct	gacctggaag	accgagaagc	ccgtgtgggt	gaaccagtgg	120
cccctgccca	agcagaagct	ggaggccctg	cacctgtctg	ccaacgagca	gctggagaag	180
ggccacatcg	agcccagctt	cagcccctgg	aacagccccg	tgctcgatgat	ccagaagaag	240
agcggcaagt	ggcgcagtgc	gaccgacctg	cgcgcctgga	acgcctgat	ccagcccatg	300
ggccccctgc	agcccggcct	gcccagcccc	gccatgatcc	ccaaggactg	gcccctgatc	360
atcatcgacc	tgaaggactg	cttcttcacc	atccccctgg	ccgagcagga	ctgcgagaag	420
ttcgcttca	ccatccccgc	catcaacaac	aaggagcccc	ccacccgctt	ccagtggaaag	480
gtgtgcccc	agggcatgct	gaacagcccc	accatctgcc	agaccttcgt	gggcccgcgc	540
ctgcagcccg	tgcgcgagaa	gttcagcgac	tgctacatca	tccactgcat	cgacgacatc	600
ctgtgcgccg	ccgagaccaa	ggacaagctg	atcgactgct	acaccttcct	gcaggccgag	660
gtggccaacg	ccggcctggc	catcgccagc	gacaagatcc	agaccagcac	ccccttccac	720
tacctgggca	tgcagatcga	gaaccgcaag	atcaagcccc	agaagatcga	gatccgcaag	780
gacaccctga	agaccctgaa	cgacttccag	aagctgctgg	gcgacatcaa	ctggatccgc	840
cccaccctgg	gcatccccac	ctacgcatat	agcaacctgt	tcagcatcct	gcgcggcgac	900
agcgactga	acagcaagcg	catgctgacc	cccaggcca	ccaaggagat	caagctgggt	960
gaggagaaga	tccagagcgc	ccagatcaac	cgcactcgacc	ccctggcccc	cctgcagctg	1020
ctgatcttcg	ccaccgcccc	cagccccacc	ggcatcatca	tccagaacac	cgacctgggt	1080
gagtggagct	tcctgccccca	cagcaccgtg	aagaccttca	ccctgtacct	ggaccagatc	1140
gccaccctga	tcggccagac	ccgcctgcgc	atcatcaagc	tgtgcggcaa	cgaccccgac	1200
aagatcgtgg	tgcccctgac	caaggagcag	gtgcgccagg	ccttcatcaa	cagcggcgcc	1260
tgaagatcgc	gcctggccaa	cttcgtgggc	atcatcgaga	accactaccc	caagaccaag	1320

Substitute Sequence Listing_USSN 10587032_PP019482.007

atcttccagt	tcctgaagct	gaccacctgg	atcctgcccc	agatcacccg	ccgcgagccc	1380
ctggagaacg	ccctgaccgt	gttcaccgac	ggcagcagca	acggcaaggc	cgccctacacc	1440
ggccccaagg	agcgcgtgat	caagaccccc	taccagagcg	cccagcgcgc	cgagctggtg	1500
gccgtgatca	ccgtgctgca	ggacttcgac	cagcccatca	acatcatcag	cgacagcgcc	1560
tacgtggtgc	aggccacccg	cgacgtggag	accgccctga	tcaagtacag	catggacgac	1620
cagctgaacc	agctgttcaa	cctgctgcag	cagaccgtgc	gcaagcgcaa	cttccccttc	1680
tacatcaccc	acatccgcgc	ccacaccaac	ctgcccggcc	ccctgaccaa	ggccaacgag	1740
caggccgacc	tgctggtgag	cagcgcctg	atcaaggccc	aggagctgca	cgccctgacc	1800
cacgtgaacg	ccgccggcct	gaagaacaag	ttcgacgtga	cctggaagca	ggccaaggac	1860
atcgtgcagc	actgcaccca	gtgccagggtg	ctgcacctgc	ccaccagga	ggccggcggtg	1920
aacccccgcg	gcctgtgccc	caacgcctg	tggcagatgg	acgtgaccca	cgtgcccagc	1980
ttcggccgcc	tgagctacgt	gcacgtgacc	gtggacacct	acagccactt	catctgggccc	2040
acctgccaga	ccggcgagag	caccagccac	gtgaagaagc	acctgctgag	ctgcttcgcc	2100
gtgatgggct	tgcccagaaa	gatcaagacc	gacaacggcc	ccggctactg	cagcaaggcc	2160
ttccagaagt	tcctgagcca	gtggaagatc	agccacacca	ccggcatccc	ctacaacagc	2220
cagggccagg	ccatcgtgga	gcgcaccaac	cgcacctga	agaccagct	ggtgaagcag	2280
aaggaggcgc	gcgacagcaa	ggagtgcacc	acccccaga	tgcagctgaa	cctggccctg	2340
tacaccctga	acttctgaa	catctaccgc	aaccagacca	ccaccagcgc	cgagcagcac	2400
ctgaccggca	agaagaacag	cccccacgag	ggcaagctga	tctggtggaa	ggacaacaag	2460
aacaagacct	gggagatcgg	caagggtgatc	acctggggcc	gcggcttcgc	ctgcgtgagc	2520
cccggcgaga	accagctgcc	cgtgtggatc	cccaccgcgc	acctgaagtt	ctacaacgag	2580
cccatccgcg	acgccaagaa	gagcaccagc	gccgagaccg	agaccagcca	gagcagcacc	2640
gtggacagcc	aggacgagca	gaacggcgac	gtgcgcgcga	ccgacgaggt	ggccatccac	2700
caggagggcc	gcgccgcaaa	cctgggcacc	accaaggagg	ccgacgccgt	gagctacaag	2760
atcagccgcg	agcacaaggg	cgacaccaac	ccccgcgagt	acgccgcctg	cagcctggac	2820
gactgcatca	acggcggcaa	gagcccctac	gcctgcccga	gcagctgcag	cgcttaa	2877

<210> 67
 <211> 106
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Manipulated cORF

<400> 67
 Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
 1 5 10 15
 His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
 20 25 30
 Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
 35 40 45
 Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
 50 55 60
 Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
 65 70 75 80
 Leu Met Ile Val Ser Met Val Ser Ala Gly Val Pro Asn Ser Ser Glu
 85 90 95
 Glu Thr Ala Thr Ile Glu Asn Gly Pro Ala
 100 105

<210> 68
 <211> 145
 <212> PRT
 <213> Artificial Sequence

<220>

<223> Manipulated PCAP5

<400> 68

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
 1 5 10 15
 His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
 20 25 30
 Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
 35 40 45
 Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
 50 55 60
 Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
 65 70 75 80
 Leu Met Ile Val Ser Met Val Val Tyr Pro Thr Ala Pro Lys Arg Gln
 85 90 95
 Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys
 100 105 110
 Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys
 115 120 125
 Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu Tyr
 130 135 140

Ala
145

<210> 69

<211> 666

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 69

Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
 1 5 10 15
 Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
 20 25 30
 Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
 35 40 45
 Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly
 50 55 60
 Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
 65 70 75 80
 Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
 85 90 95
 Thr Glu Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Ile
 100 105 110
 Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
 115 120 125

Substitute Sequence Listing_USSN 10587032_PP019482.007

Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
130 135 140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145 150 155 160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
165 170 175

Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
180 185 190

Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
195 200 205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
210 215 220

Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
225 230 235 240

Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
245 250 255

Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
260 265 270

Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
275 280 285

Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
290 295 300

Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
305 310 315 320

Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
325 330 335

Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
340 345 350

Leu Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
355 360 365

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
370 375 380

Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln
385 390 395 400

Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
405 410 415

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Glu Lys Ile Gln Asp
420 425 430

Pro Gly Ser Thr Cys Pro Ser Phe Asn Thr Val Arg Gln Gly Ser Lys
435 440 445

Glu Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Val Ala Gln Lys
450 455 460

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ser Ile Ala Asp Glu Lys Ala Arg Lys Val Ile Val Glu Leu Met Ala
 465 470 475 480

Tyr Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys
 485 490 495

Gly Lys Val Pro Ala Gly Ser Asp Val Ile Ser Glu Tyr Val Lys Ala
 500 505 510

Cys Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln
 515 520 525

Ala Ile Thr Gly Val Val Leu Gly Gly Gln Val Arg Thr Phe Gly Arg
 530 535 540

Lys Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Lys Asn Cys Pro
 545 550 555 560

Val Leu Asn Lys Gln Asn Ile Thr Ile Gln Ala Thr Thr Thr Gly Arg
 565 570 575

Glu Pro Pro Asp Leu Cys Pro Arg Cys Lys Lys Gly Lys His Trp Ala
 580 585 590

Ser Gln Cys Arg Ser Lys Phe Asp Lys Asn Gly Gln Pro Leu Ser Gly
 595 600 605

Asn Glu Gln Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
 610 615 620

Pro Ile Gln Pro Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Pro
 625 630 635 640

Leu Ser Gln Val Phe Gln Gly Ile Ser Gln Leu Pro Gln Tyr Asn Asn
 645 650 655

Cys Pro Pro Pro Gln Ala Ala Val Gln Gln
 660 665

<210> 70
 <211> 667
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Manipulated Gag

<400> 70
 Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
 1 5 10 15

Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
 20 25 30

Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
 35 40 45

Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly
 50 55 60

Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
 65 70 75 80

Substitute Sequence Listing_USSN 10587032_PP019482.007

Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
85 90 95

Thr Glu Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Ile
100 105 110

Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
115 120 125

Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
130 135 140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145 150 155 160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
165 170 175

Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
180 185 190

Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
195 200 205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
210 215 220

Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
225 230 235 240

Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
245 250 255

Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
260 265 270

Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
275 280 285

Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
290 295 300

Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
305 310 315 320

Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
325 330 335

Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
340 345 350

Leu Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
355 360 365

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
370 375 380

Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln
385 390 395 400

Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
405 410 415

Substitute Sequence Listing_USSN 10587032_PP019482.007

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Glu Lys Ile Gln Asp
420 425 430

Pro Gly Ser Thr Cys Pro Ser Phe Asn Thr Val Arg Gln Gly Ser Lys
435 440 445

Glu Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Val Ala Gln Lys
450 455 460

Ser Ile Ala Asp Glu Lys Ala Arg Lys Val Ile Val Glu Leu Met Ala
465 470 475 480

Tyr Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys
485 490 495

Gly Lys Val Pro Ala Gly Ser Asp Val Ile Ser Glu Tyr Val Lys Ala
500 505 510

Cys Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln
515 520 525

Ala Ile Thr Gly Val Val Leu Gly Gly Gln Val Arg Thr Phe Gly Arg
530 535 540

Lys Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Lys Asn Cys Pro
545 550 555 560

Val Leu Asn Lys Gln Asn Ile Thr Ile Gln Ala Thr Thr Thr Gly Arg
565 570 575

Glu Pro Pro Asp Leu Cys Pro Arg Cys Lys Lys Gly Lys His Trp Ala
580 585 590

Ser Gln Cys Arg Ser Lys Phe Asp Lys Asn Gly Gln Pro Leu Ser Gly
595 600 605

Asn Glu Gln Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
610 615 620

Pro Ile Gln Pro Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Pro
625 630 635 640

Leu Ser Gln Val Phe Gln Gly Ile Ser Gln Leu Pro Gln Tyr Asn Asn
645 650 655

Cys Pro Pro Pro Gln Ala Ala Val Gln Gln Ala
660 665

<210> 71

<211> 334

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 71

Met Trp Ala Thr Ile Val Gly Lys Arg Ala Lys Gly Pro Ala Ser Gly
1 5 10 15

Pro Thr Thr Asn Trp Gly Ile Pro Asn Ser Ala Ile Cys Ser Ser Gly
20 25 30

Phe Ser Gly Thr Thr Thr Pro Thr Val Pro Ser Val Ser Gly Asn Lys
35 40 45

Substitute Sequence Listing_USSN 10587032_PP019482.007

Pro Val Thr Thr Ile Gln Gln Leu Ser Pro Ala Thr Ser Gly Ser Ala
50 55 60

Ala Val Asp Leu Cys Thr Ile Gln Ala Val Ser Leu Leu Pro Gly Glu
65 70 75 80

Pro Pro Gln Lys Thr Pro Thr Gly Val Tyr Gly Pro Leu Pro Lys Gly
85 90 95

Thr Val Gly Leu Ile Leu Gly Arg Ser Ser Leu Asn Leu Lys Gly Val
100 105 110

Gln Ile His Thr Ser Val Val Asp Ser Asp Tyr Lys Gly Glu Ile Gln
115 120 125

Leu Val Ile Ser Ser Ser Ile Pro Trp Ser Ala Ser Pro Arg Asp Arg
130 135 140

Ile Ala Gln Leu Leu Leu Leu Pro Tyr Ile Lys Gly Gly Asn Ser Glu
145 150 155 160

Ile Lys Arg Ile Gly Gly Leu Gly Ser Thr Asp Pro Thr Gly Lys Ala
165 170 175

Ala Tyr Trp Ala Ser Gln Val Ser Glu Asn Arg Pro Val Cys Lys Ala
180 185 190

Ile Ile Gln Gly Lys Gln Phe Glu Gly Leu Val Asp Thr Gly Ala Asp
195 200 205

Val Ser Ile Ile Ala Leu Asn Gln Trp Pro Lys Asn Trp Pro Lys Gln
210 215 220

Lys Ala Val Thr Gly Leu Val Gly Ile Gly Thr Ala Ser Glu Val Tyr
225 230 235 240

Gln Ser Thr Glu Ile Leu His Cys Leu Gly Pro Asp Asn Gln Glu Ser
245 250 255

Thr Val Gln Pro Met Ile Thr Ser Ile Pro Leu Asn Leu Trp Gly Arg
260 265 270

Asp Leu Leu Gln Gln Trp Gly Ala Glu Ile Thr Met Pro Ala Pro Ser
275 280 285

Tyr Ser Pro Thr Ser Gln Lys Ile Met Thr Lys Met Gly Tyr Ile Pro
290 295 300

Gly Lys Gly Leu Gly Lys Asn Glu Asp Gly Ile Lys Ile Pro Val Glu
305 310 315 320

Ala Lys Ile Asn Gln Glu Arg Glu Gly Ile Gly Asn Pro Cys
325 330

<210> 72
<211> 335
<212> PRT
<213> Artificial Sequence

<220>
<223> Manipulated PRT

<400> 72

Substitute Sequence Listing_USSN 10587032_PP019482.007

Met	Trp	Ala	Thr	Ile	Val	Gly	Lys	Arg	Ala	Lys	Gly	Pro	Ala	Ser	Gly	1	5	10	15
Pro	Thr	Thr	Asn	Trp	Gly	Ile	Pro	Asn	Ser	Ala	Ile	Cys	Ser	Ser	Gly	20	25	30	
Phe	Ser	Gly	Thr	Thr	Thr	Pro	Thr	Val	Pro	Ser	Val	Ser	Gly	Asn	Lys	35	40	45	
Pro	Val	Thr	Thr	Ile	Gln	Gln	Leu	Ser	Pro	Ala	Thr	Ser	Gly	Ser	Ala	50	55	60	
Ala	Val	Asp	Leu	Cys	Thr	Ile	Gln	Ala	Val	Ser	Leu	Leu	Pro	Gly	Glu	65	70	75	80
Pro	Pro	Gln	Lys	Thr	Pro	Thr	Gly	Val	Tyr	Gly	Pro	Leu	Pro	Lys	Gly	85	90	95	
Thr	Val	Gly	Leu	Ile	Leu	Gly	Arg	Ser	Ser	Leu	Asn	Leu	Lys	Gly	Val	100	105	110	
Gln	Ile	His	Thr	Ser	Val	Val	Asp	Ser	Asp	Tyr	Lys	Gly	Glu	Ile	Gln	115	120	125	
Leu	Val	Ile	Ser	Ser	Ser	Ile	Pro	Trp	Ser	Ala	Ser	Pro	Arg	Asp	Arg	130	135	140	
Ile	Ala	Gln	Leu	Leu	Leu	Leu	Pro	Tyr	Ile	Lys	Gly	Gly	Asn	Ser	Glu	145	150	155	160
Ile	Lys	Arg	Ile	Gly	Gly	Leu	Gly	Ser	Thr	Asp	Pro	Thr	Gly	Lys	Ala	165	170	175	
Ala	Tyr	Trp	Ala	Ser	Gln	Val	Ser	Glu	Asn	Arg	Pro	Val	Cys	Lys	Ala	180	185	190	
Ile	Ile	Gln	Gly	Lys	Gln	Phe	Glu	Gly	Leu	Val	Asp	Thr	Gly	Ala	Asp	195	200	205	
Val	Ser	Ile	Ile	Ala	Leu	Asn	Gln	Trp	Pro	Lys	Asn	Trp	Pro	Lys	Gln	210	215	220	
Lys	Ala	Val	Thr	Gly	Leu	Val	Gly	Ile	Gly	Thr	Ala	Ser	Glu	Val	Tyr	225	230	235	240
Gln	Ser	Thr	Glu	Ile	Leu	His	Cys	Leu	Gly	Pro	Asp	Asn	Gln	Glu	Ser	245	250	255	
Thr	Val	Gln	Pro	Met	Ile	Thr	Ser	Ile	Pro	Leu	Asn	Leu	Trp	Gly	Arg	260	265	270	
Asp	Leu	Leu	Gln	Gln	Trp	Gly	Ala	Glu	Ile	Thr	Met	Pro	Ala	Pro	Ser	275	280	285	
Tyr	Ser	Pro	Thr	Ser	Gln	Lys	Ile	Met	Thr	Lys	Met	Gly	Tyr	Ile	Pro	290	295	300	
Gly	Lys	Gly	Leu	Gly	Lys	Asn	Glu	Asp	Gly	Ile	Lys	Ile	Pro	Val	Glu	305	310	315	320
Ala	Lys	Ile	Asn	Gln	Glu	Arg	Glu	Gly	Ile	Gly	Asn	Pro	Cys	Ala		325	330	335	

Substitute Sequence Listing_USSN 10587032_PP019482.007

<210> 73
 <211> 957
 <212> PRT
 <213> Human endogenous retrovirus, K family (HERV-K)

 <400> 73
 Met Asn Lys Ser Arg Lys Arg Arg Asn Arg Glu Ser Leu Leu Gly Ala
 1 5 10 15
 Ala Thr Val Glu Pro Pro Lys Pro Ile Pro Leu Thr Trp Lys Thr Glu
 20 25 30
 Lys Pro Val Trp Val Asn Gln Trp Pro Leu Pro Lys Gln Lys Leu Glu
 35 40 45
 Ala Leu His Leu Leu Ala Asn Glu Gln Leu Glu Lys Gly His Ile Glu
 50 55 60
 Pro Ser Phe Ser Pro Trp Asn Ser Pro Val Phe Val Ile Gln Lys Lys
 65 70 75 80
 Ser Gly Lys Trp Arg Met Leu Thr Asp Leu Arg Ala Val Asn Ala Val
 85 90 95
 Ile Gln Pro Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met
 100 105 110
 Ile Pro Lys Asp Trp Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe
 115 120 125
 Phe Thr Ile Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr
 130 135 140
 Ile Pro Ala Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys
 145 150 155 160
 Val Leu Pro Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe
 165 170 175
 Val Gly Arg Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr
 180 185 190
 Ile Ile His Cys Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp
 195 200 205
 Lys Leu Ile Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala
 210 215 220
 Gly Leu Ala Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His
 225 230 235 240
 Tyr Leu Gly Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile
 245 250 255
 Glu Ile Arg Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu
 260 265 270
 Leu Gly Asp Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr
 275 280 285
 Ala Met Ser Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn
 290 295 300

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ser Lys Arg Met Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val
 305 310 315 320
 Glu Glu Lys Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala
 325 330 335
 Pro Leu Gln Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile
 340 345 350
 Ile Ile Gln Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser
 355 360 365
 Thr Val Lys Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile
 370 375 380
 Gly Gln Thr Arg Leu Arg Ile Ile Lys Leu Cys Gly Asn Asp Pro Asp
 385 390 395 400
 Lys Ile Val Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile
 405 410 415
 Asn Ser Gly Ala Trp Lys Ile Gly Leu Ala Asn Phe Val Gly Ile Ile
 420 425 430
 Asp Asn His Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr
 435 440 445
 Thr Trp Ile Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala
 450 455 460
 Leu Thr Val Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr
 465 470 475 480
 Gly Pro Lys Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg
 485 490 495
 Ala Glu Leu Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro
 500 505 510
 Ile Asn Ile Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp
 515 520 525
 Val Glu Thr Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln
 530 535 540
 Leu Phe Asn Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe
 545 550 555 560
 Tyr Ile Thr His Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr
 565 570 575
 Lys Ala Asn Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys
 580 585 590
 Ala Gln Glu Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys
 595 600 605
 Asn Lys Phe Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His
 610 615 620
 Cys Thr Gln Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val
 625 630 635 640

Substitute Sequence Listing_USSN 10587032_PP019482.007

```

Asn Pro Arg Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr
645                               650                               655

His Val Pro Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp
660                               665                               670

Thr Tyr Ser His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr
675                               680                               685

Ser His Val Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val
690                               695                               700

Pro Glu Lys Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala
705                               710                               715                               720

Phe Gln Lys Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile
725                               730                               735

Pro Tyr Asn Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr
740                               745                               750

Leu Lys Thr Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu
755                               760                               765

Cys Thr Thr Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn
770                               775                               780

Phe Leu Asn Ile Tyr Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His
785                               790                               795                               800

Leu Thr Gly Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp
805                               810                               815

Lys Asp Asn Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp
820                               825                               830

Gly Arg Gly Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val
835                               840                               845

Trp Ile Pro Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Arg Asp
850                               855                               860

Ala Lys Lys Ser Thr Ser Ala Glu Thr Glu Thr Ser Gln Ser Ser Thr
865                               870                               875                               880

Val Asp Ser Gln Asp Glu Gln Asn Gly Asp Val Arg Arg Thr Asp Glu
885                               890                               895

Val Ala Ile His Gln Glu Gly Arg Ala Ala Asn Leu Gly Thr Thr Lys
900                               905                               910

Glu Ala Asp Ala Val Ser Tyr Lys Ile Ser Arg Glu His Lys Gly Asp
915                               920                               925

Thr Asn Pro Arg Glu Tyr Ala Ala Cys Ser Leu Asp Asp Cys Ile Asn
930                               935                               940

Gly Gly Lys Ser Pro Tyr Ala Cys Arg Ser Ser Cys Ser
945                               950                               955

```

<210> 74
 <211> 958
 <212> PRT

<213> Artificial Sequence

<220>

<223> Manipulated Pol

<400> 74

Met Asn Lys Ser Arg Lys Arg Arg Asn Arg Glu Ser Leu Leu Gly Ala
 1 5 10 15
 Ala Thr Val Glu Pro Pro Lys Pro Ile Pro Leu Thr Trp Lys Thr Glu
 20 25 30
 Lys Pro Val Trp Val Asn Gln Trp Pro Leu Pro Lys Gln Lys Leu Glu
 35 40 45
 Ala Leu His Leu Leu Ala Asn Glu Gln Leu Glu Lys Gly His Ile Glu
 50 55 60
 Pro Ser Phe Ser Pro Trp Asn Ser Pro Val Phe Val Ile Gln Lys Lys
 65 70 75 80
 Ser Gly Lys Trp Arg Met Leu Thr Asp Leu Arg Ala Val Asn Ala Val
 85 90 95
 Ile Gln Pro Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met
 100 105 110
 Ile Pro Lys Asp Trp Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe
 115 120 125
 Phe Thr Ile Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr
 130 135 140
 Ile Pro Ala Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys
 145 150 155 160
 Val Leu Pro Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe
 165 170 175
 Val Gly Arg Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr
 180 185 190
 Ile Ile His Cys Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp
 195 200 205
 Lys Leu Ile Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala
 210 215 220
 Gly Leu Ala Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His
 225 230 235 240
 Tyr Leu Gly Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile
 245 250 255
 Glu Ile Arg Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu
 260 265 270
 Leu Gly Asp Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr
 275 280 285
 Ala Met Ser Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn
 290 295 300

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ser Lys Arg Met Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val
305 310 315 320

Glu Glu Lys Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala
325 330 335

Pro Leu Gln Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile
340 345 350

Ile Ile Gln Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser
355 360 365

Thr Val Lys Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile
370 375 380

Gly Gln Thr Arg Leu Arg Ile Ile Lys Leu Cys Gly Asn Asp Pro Asp
385 390 395 400

Lys Ile Val Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile
405 410 415

Asn Ser Gly Ala Trp Lys Ile Gly Leu Ala Asn Phe Val Gly Ile Ile
420 425 430

Asp Asn His Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr
435 440 445

Thr Trp Ile Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala
450 455 460

Leu Thr Val Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr
465 470 475 480

Gly Pro Lys Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg
485 490 495

Ala Glu Leu Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro
500 505 510

Ile Asn Ile Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp
515 520 525

Val Glu Thr Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln
530 535 540

Leu Phe Asn Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe
545 550 555 560

Tyr Ile Thr His Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr
565 570 575

Lys Ala Asn Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys
580 585 590

Ala Gln Glu Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys
595 600 605

Asn Lys Phe Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His
610 615 620

Cys Thr Gln Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val
625 630 635 640

Substitute Sequence Listing_USSN 10587032_PP019482.007

Asn Pro Arg Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr
645 650 655

His Val Pro Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp
660 665 670

Thr Tyr Ser His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr
675 680 685

Ser His Val Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val
690 695 700

Pro Glu Lys Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala
705 710 715 720

Phe Gln Lys Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile
725 730 735

Pro Tyr Asn Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr
740 745 750

Leu Lys Thr Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu
755 760 765

Cys Thr Thr Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn
770 775 780

Phe Leu Asn Ile Tyr Arg Asn Gln Thr Thr Thr Ser Ala Glu Gln His
785 790 795 800

Leu Thr Gly Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp
805 810 815

Lys Asp Asn Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp
820 825 830

Gly Arg Gly Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val
835 840 845

Trp Ile Pro Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Arg Asp
850 855 860

Ala Lys Lys Ser Thr Ser Ala Glu Thr Glu Thr Ser Gln Ser Ser Thr
865 870 875 880

Val Asp Ser Gln Asp Glu Gln Asn Gly Asp Val Arg Arg Thr Asp Glu
885 890 895

Val Ala Ile His Gln Glu Gly Arg Ala Ala Asn Leu Gly Thr Thr Lys
900 905 910

Glu Ala Asp Ala Val Ser Tyr Lys Ile Ser Arg Glu His Lys Gly Asp
915 920 925

Thr Asn Pro Arg Glu Tyr Ala Ala Cys Ser Leu Asp Asp Cys Ile Asn
930 935 940

Gly Gly Lys Ser Pro Tyr Ala Cys Arg Ser Ser Cys Ser Ala
945 950 955

<210> 75
<211> 12366
<212> DNA

Substitute Sequence Listing_USSN 10587032_PP019482.007

<213> Human endogenous retrovirus, K family (HERV-K), located at 22q11.2

<400> 75

tgtggggaaa	agaaagagag	atcagactgt	tactgtgtct	atgtagaaag	aaatagacat	60
aagagactcc	attttgttct	gtactaagaa	aaattcttct	gctttgagat	gctgttaatc	120
tgtaaccta	gccccaaccc	tgtgtctaca	gaaacaggtg	ctgtgttgac	tcaaggttta	180
atggattcag	ggctgtgcag	gatgtgcttt	gttaaaca	tgcttgaagg	cagcaagctt	240
gttaagagtc	atcaccactc	cctaattctca	agtaagcagg	gacacaaaca	ctgcggaagg	300
ccgcagggac	ctctgcctag	gaaagccagg	tgttgtccaa	ggtttctccc	catgtgacag	360
tctgaaatat	ggcctcttgg	gaagggaaag	acctgactgt	cccctggccc	gacacccgta	420
aagggtctgt	gctgaggatt	agtaaaagag	gaaggaaggc	ctctttgcag	ttgagataag	480
aggaaggcat	ctgtctcctg	ctcatccctg	ggcaatggaa	tgtcttggtg	taaagcctga	540
ttgtatatgc	catctactga	gataggagaa	aactgcctta	gggctggagg	tgggacatgc	600
tggcggcaat	actgtctctt	aaggcattga	gatgtttatg	tatatgcaca	tcaaaagcac	660
agcacctttt	tctttacctt	gtttatgatg	cagagacatt	tgttcacatg	ttttcctgct	720
ggccctctcc	ccactattac	cctattgtcc	tgccacatcc	ccctctccga	gatggtagag	780
ataatgatca	ataaatactg	agggaaactca	gagaccggtg	cggcgcgggt	cctccatatg	840
ctgagcgccg	gtccccctggg	cccacttttc	tttctctata	ctttgtctct	gttgtctttc	900
ttttctcaag	tctctcgttc	cacctgagga	gaaatgccca	cagctgtgga	ggcgcaggcc	960
actccatctg	gtgcccacg	tggtatgctt	tctctagggg	gaagggactc	tcgagtgtgg	1020
tcattgagga	caagtcaacg	agagattccc	gagtacgtct	acagtgagcc	ttgtggtaag	1080
cttgggcgct	cggaagaagc	cagggttaat	ggggcaact	aaaagtaaag	tctctcattc	1140
cacctgatga	gaaacaccca	gaggtgtgga	ggggcaggcc	accccttcag	ggtaggggtcc	1200
cctccatgca	gaccatagag	cacaggtgtg	ccccaaagag	gagcagagag	aaggaggag	1260
agggccacg	agagacttgg	aatgaatgg	caggatttta	ggcgtggac	ttgggttcgg	1320
ggcacctggc	ctttccttgt	gtatttctcc	tactgtctgc	ctaactattt	aatacaataa	1380
aagaaaacca	gccccctggt	cttgtgggtg	ttccaccttc	ccgggtcccc	gctggctgcc	1440
tggcttcctc	ccgcagctcc	tgctgtgtgt	gtatgtgtgt	gtgtgtgcac	atctgtgggg	1500
cgtatgtgtg	ttcgtctttg	taattgaggc	tgtagagtgg	agagagcagg	ggttttctct	1560
ggggacccag	agagaaggag	gcgttttcac	cacagccgaa	cagggcagga	ccccagcacc	1620
cgggacccag	cgggactttg	ccaaggggat	ggacctggct	gggccacgcg	gctgttttgt	1680
tagggaaaa	aaagagagat	cacactgtta	ctgtgtctat	gtagaaaagg	aagacataaa	1740
ctccattttg	agctgtacta	agaaaaatta	ttttgccttg	acctgctgtt	aacctgtaac	1800
tgtagcccca	accctgtgct	caaagaaaca	tgtgtgtgat	ggaatcaagg	tttaagggat	1860
caagggtgtg	acaggatgtg	ccttggttaac	aatgtgttta	caggcagtat	gcttggtaaa	1920
agtcacgccc	attctccatt	ctccattaat	caggggcacg	atgcactgcg	gaaagccaca	1980
gggacctctg	cccagaaaag	cctgggtatt	gtccaaggct	tccccccact	gagacagcct	2040
gagatacggc	ctcgtgggaa	gggaaagacc	tgaccgtccc	ccagcccagc	acccgtaaaag	2100
ggctctgtcg	gaggaggatt	agtaaaaggg	gaagccctct	tgtagttgag	ataagaggaa	2160
ggcctccgtc	tcctgcatgt	ccttggggat	ggaatgtctt	gggtgaaaac	ccgatagtac	2220
attccttcta	ttctgagaga	agaaaaccac	cctgtggctg	gaggtgagat	atgctagcgg	2280
caatgctgct	ctgttactct	ttgctacact	gagatgtttg	gggtggagaga	agcataaatc	2340
tggcctatgt	gcacatctgg	gcacagaacc	tccccttgaa	cttgtgacac	agattccttt	2400
gttcacatgt	tttcctgtgt	accttctccc	cactatcgcc	ctgttctccc	accgcattcc	2460
ccttgctgag	atagtgtaaa	tagtaatctg	tagataccaa	gggaactcag	agaccatggc	2520
cgggtgcacat	cctccgtacg	ctgagcgtg	gtcccctggg	cccattgttc	tttctctata	2580
ctttgtctct	gtgtcttatt	tctttcctca	gtctctcatc	cctcctgacg	agaaataccc	2640
acagggtgtg	aggggctggc	ccccttcac	tgatgcccaa	tgtgggtgcc	tttctctagg	2700
gtgaagggtac	tctacagtgt	ggtcattgag	gacaagttga	cgagagagtc	ccaagtacgt	2760
ccacggtcag	ccttgcggtg	agcttgtgtg	cttagaggaa	cccagggtaa	cgatggggca	2820
aactgaaaag	aaatatgcct	cttatctcag	ctttattaaa	attcttttaa	gaagaggggg	2880
agttagagct	tctacagaaa	atctaattac	gctattttcaa	acaatagaac	aattctgccc	2940
atggtttcca	gaacagggaa	ctttagatct	aaaagattgg	gaaaaaattg	gcaaagaatt	3000
aaaacaagca	aatagggaag	gtaaaatcat	cccacttaca	gtatggaatg	attgggccat	3060
tattaaagca	actttagaac	catttcaaac	aggagaagat	attgtttcag	tttctgatgc	3120
ccctaaaagc	tgtgtaacag	attgtgaaga	agaggcaggg	acagaatccc	agcaagggaac	3180
ggaaagtcca	cattgtaaat	atgtagcaga	gtctgtaatg	gctcagtcac	cgcaaaatgt	3240
tgactacagt	caattacagg	agataaatata	ccctgaaatca	tcaaaattgg	gggaaggagg	3300
tccagaatca	ttggggccat	cagagcctaa	accacgatcg	ccatcaactc	ctcctcccgt	3360
ggttcagatg	cctgtaacat	tacaacctca	aacgcagggt	agacaagcac	aaaccccaag	3420
agaaaatcaa	gtagaaaagg	acagagtcct	tatcccggca	atgccaaact	agatacagta	3480
tccacaatat	cagccggtag	aaaataagac	ccaaccgctg	gtagtttatc	aataccggct	3540
gccaaccgag	cttcagtatc	ggcctccttc	agaggttcaa	tacagacctc	aagcgggtgtg	3600

Substitute Sequence Listing_USSN 10587032_PP019482.007

tcctgtgcc	aatagcacg	caccatacca	gcaaccaca	gcgatggcg	ctaattcacc	3660
agcaacacag	gacgcggcg	tgtatcctca	gccgcccact	gtgagactta	atcctacagc	3720
atcacgtagt	ggacaggggtg	gtgcactgca	tgcagtcatt	gatgaagcca	gaaaacaggg	3780
cgatcttgag	gcatggcggt	tcctggtaat	tttacaactg	gtacaggccg	gggaagagac	3840
tcaagtagga	gcgcctgccc	gagctgagac	tagatgtgaa	cctttcacca	tgaaaatggt	3900
aaaagatata	aaggaaggag	ttaaacaata	tggatccaac	tccccttata	taagaacatt	3960
attagattcc	attgctcatg	gaaatagact	tactccttat	gactgggaaa	ttttggccaa	4020
atcttccctt	tcatcctctc	agtatctaca	gtttaaaaacc	tggtggattg	atggagtaca	4080
agaacaggta	cgaaaaaatc	aggctactaa	gccactgtt	aatatagacg	cagaccaatt	4140
gttaggaaca	ggtccaaatt	ggagcaccat	taaccaacaa	tcagtgatgc	agaatgaggc	4200
tattgaacaa	gtaagggcta	tttgccctcag	ggcctgggga	aaaattcagg	acccaggaac	4260
agctttccct	attaattcaa	ttagacaagg	ctctaaagag	ccatatcctg	actttgtggc	4320
aagattacaa	gatgctgctc	aaaagtctat	tacagatgac	aatgcccga	aagttattgt	4380
agaattaatg	gcctatgaaa	atgcaaatcc	agaatgtcag	tcggccataa	agccattaaa	4440
aggaaaagtt	ccagcaggag	ttgatgtaat	tacagaatat	gtgaaggctt	gtgatgggat	4500
tggaggagct	atgcataagg	caatgcta	ggctcaagca	atgagggggc	tcactctagg	4560
aggacaagtt	agaacatttg	ggaaaaaatg	ttataattgt	ggtcaaatac	gtcatctgaa	4620
aaggagttgc	ccaggcttaa	ataaacagaa	tataataaat	caagctatta	cagcaaaaaa	4680
taaaaagcca	tctggcctgt	gtccaaaatg	tggaaaagca	aaacattggg	ccaatcaatg	4740
tcattctaaa	tttgataaag	atgggcaacc	atttgtctgga	aacaggaaga	ggggccagcc	4800
tcaggccccc	caacaaactg	gggcattccc	agtttaactg	tttgttcctc	aggggtttca	4860
aggacaacaa	cccctacaga	aaataccacc	acttcaggga	gtcagccaat	tacaacaatc	4920
caacagctgt	cccgcgccac	agcaggcagc	accgcagtag	atttatgttc	cacccaaatg	4980
gtctttttac	tccctggaaa	gccccacaa	aagattccta	gaggggtata	tggcccgtctg	5040
ccagaaggga	gggtaggcct	ttgagggaga	tcaagtctaa	atttgaaggg	agtccaaatt	5100
catactgggg	taatttattc	agattataaa	gggggaattc	agttagtgat	cagctccact	5160
gttccccgga	gtgccaatcc	aggatgata	attgtcta	tactgctttt	gccttatggt	5220
aaaattgggg	aaaacaaaaa	ggaaagaaca	ggagggtttg	gaagtaccaa	ccctgcagga	5280
aaagctgctt	attgggctaa	tcagggtctca	gaggatagac	ccgtgtgtac	agtcactatt	5340
cagggaaaga	gtttgaagga	ttagtggata	cccaggctga	tgtttctgtc	atcggcatag	5400
gtactgcctc	agaagtgtat	caaagtgcc	tgattttata	ttgtccagga	tctgataatc	5460
aagaaagtag	ggttcagcct	gtgatcactt	cattccaatc	aatttatggg	gccgagactt	5520
gttacaacaa	tggcatgag	agattactat	cccagctccc	ctatacagcc	ccaggaataa	5580
aaaaatcatg	actaaaatgg	gatagctccc	taaaaaggga	ctaggaaaga	agtcccaatt	5640
gaggctgaaa	aaaatcaaaa	aagaaaagga	atagggcatc	cttttttagga	gcggtcactg	5700
tagagcctcc	aaaaccctatt	ccattaactt	gggggaaaaa	aaaacaactg	tatggtaaat	5760
cagcagcgct	tccaaaacaa	aaactggagg	ctttacattt	attagcaaag	aaacaattag	5820
aaaaaggaca	ttgagccttc	attttcgctt	tgggaattctg	tttgtaattc	agaaaaaatc	5880
cggcagatgg	cgtataatgc	cgtaattcaa	cccattggggg	ctctcccacc	ccgggtggccc	5940
tctccagcca	tggtcccttt	taattataat	tgatctgaag	gattgctttt	ttaccattcc	6000
tctggcaaaa	caggattttg	aaaaatttgc	ttttaccaca	ccagcctaaa	taataaagaa	6060
ccagccacca	ggtttcagtg	gaaagtattg	cctcagggaa	tgcttaatat	ttcaactatt	6120
gtcagctca	agctctgcaa	ccagttagag	acaagttttc	agactgttac	atcgttcact	6180
atgttgatat	tttgtgtgct	gcagaaacga	gagacaattt	aattgaccgt	tacacatttc	6240
tgagacagag	ggttgccaac	gcgggactga	caataacatc	tgataagatt	caaacctcta	6300
ctcctttccg	ttacttggga	atgcaggtag	aggaaaggaa	aattaaacca	caaaaaatag	6360
aaataagaaa	agacacatta	aaagcattaa	atgagtttca	aaagttgcta	ggagatacta	6420
attggatttg	gagatattaa	ttggattttg	ccaactctag	gcattcctac	ttatgccatg	6480
tcaaatttgt	tctctttctt	aagaggggac	tcggaattaa	atagtgaag	aacgttaact	6540
ccagaggcaa	ctaaagaaat	taaatttaatt	gaagaaaaaa	ttcggtcagc	acaagtaaat	6600
agaatagatc	acttggcccc	actccaaatt	ttgatttttg	ctactgcaca	ttccctaaca	6660
ggcatcattg	ttcaaaaatac	agatcttgtg	gagtgttcct	tccttctca	cagtacaatt	6720
aagactttta	cattgtactt	ggatcaaatg	gctacattaa	ttggtcaggg	aagattatga	6780
ataataacat	tgtgtgga	tgaccagat	aaaatcactg	ttcctttcaa	caagcaacag	6840
gtagacaag	cctttatcaa	ttctgggtga	tggcagattg	gtcttgccga	ttttgtggga	6900
attattgaca	atcgttacc	caaaaacaaa	atcttccagt	ttttaaaatt	gactacttgg	6960
attttaccta	aagttacc	acataagcct	tttaaaaaatg	ctctggcagt	gtttactgat	7020
ggttccagca	atggaaaagt	ggcttacacc	gggccaaaag	aatgagtc	caaaactcag	7080
tatcacttga	ctcaaagagc	agagttgggt	gccgtcatta	cagtgttaac	aagattttta	7140
tcagtctatt	aacattgtat	cagattctgc	atatgtagta	caggctacaa	aggatattga	7200
gagagcccta	atcaaataca	ttatggatga	tcagttaa	ccgctgttta	atttgtttaca	7260
acaaaatgta	agaaaaagaa	atttcccatt	ttatattact	catattcgag	cacacactaa	7320
tttaccaggg	ccittaaacta	aagcaaatga	acaagctgac	ttgctagt	catctgcatt	7380

Substitute Sequence Listing_USSN 10587032_PP019482.007

catggaagca	caagaacttc	atgccttgac	tcatgtaaat	gcaataggat	taaaaaataa	7440
atttgatata	acatggaaac	agacaaaaaa	tattgtacaa	cattgcaccc	agtgtcagat	7500
tctacacctg	gccactcagg	aggcaagagt	taatccaga	ggtctatgtc	ctaattgtgt	7560
atggcaaagt	gatgtcatgc	acgtaccttc	atttggaaaa	ttgtcatttg	tccatgtgac	7620
agttgatact	tattcacatt	tcatatgggc	aacctgccag	acaggagaaa	gtacttccca	7680
tgttaaaaga	cattttattat	cttgttttcc	tgatcatggga	gttccagaaa	aagttaaaac	7740
agacaatggg	ccaggttact	gtagtaaagc	agttcaaaaa	ttcttaaata	agtggaaaat	7800
tacacataca	ataggaattc	tctataattc	ccaaggacag	gccataattg	aaagaactaa	7860
tagaacactc	aaagctcaat	tggttaaaca	aaaaaaagga	aaagacagga	gtataacact	7920
ccccagatgc	aacttaattct	agcactctat	actttaaatg	ttttaaacat	ttatagaaat	7980
cagaccacta	cctctgcaga	acaacatctt	actggtaaaa	ggaacagccc	acatgaagga	8040
aaactgattt	ggtggaaaga	taataaaaaat	aaaacatggg	aaatggggaa	ggtgataacg	8100
tggtgggagag	gttttgcttg	tgtttcacca	ggagaaaatc	agcttcctgt	ttggataccc	8160
actagacatt	taaagttcta	caatgaactc	actggagatg	caaagaaaag	tgtggagatg	8220
gagacacccc	aatcgactcg	ccaggtaaac	aaaatggatg	tatcagaaga	acagaaaaag	8280
ttgccttcca	tcaaggaagc	agagttgcca	atataggcac	aattaaagaa	gctgacacag	8340
ttagctaaaa	aaaaaagcct	agagaataca	aaggtgacac	caactccaga	gaatatgctg	8400
cttgacagctc	tgatgattgt	atcaacgggtg	gtaagtcttc	ccaagtctgc	aggagcagct	8460
gcagctaatt	atacttactg	ggcctatgtg	cctttcccac	ccttaattcg	ggcagttaca	8520
tagatggata	atcctattga	agtagatgtt	aataatagtg	catgggtgcc	tggccccaca	8580
gatgactggt	gccctggcca	acctgaagaa	ggaatgatga	tgaatatattc	cattgggtat	8640
ccttatcctc	ctgtttgcct	agggaaaggca	ccaggatgct	taatgcctac	aacccaaaat	8700
tggttggtag	aagtacctac	agtcagtgtc	accagtagat	ttacttatca	catggtaagt	8760
ggaatgtcac	agataaataa	tttacaggac	ccttcttatc	aaagatcatt	acaatgtagg	8820
cctaagggga	aggcttgccc	caaggaaatt	cccaaagaat	caaaaagccc	agaagtctta	8880
gtctgcggag	aatgtgtggc	tgatactgca	gtgtagtaca	aaacaatgaa	ttttgaacta	8940
tgatagactg	ggtcccctga	ggccaattat	atcataactg	tacaggccag	actcattcat	9000
gttcacaggc	cccattccatc	tggtccatta	atccagccta	tgacggtgat	gtaactgaaa	9060
ggctggacca	ggtttataga	aggttagaat	cactctgtcc	aaggaaatgg	ggtgaaaagg	9120
gaatttcattc	accttgacca	aagttagtcc	tgttactggt	cctgaacatc	cagaattagg	9180
aagcttactg	tggcctcaca	ccacattaga	atttgttctg	gaaatcaagc	tataggaaca	9240
agagatcgta	agtcataatta	tactatcaac	ctaaattcca	gtctgacaat	tcctttgcaa	9300
aattgtgtaa	aactccctta	tattgctagt	tgtaggaaaa	acatagtatt	taaacctgat	9360
tcccaaacca	taatctgtga	aaattgtgga	atgtttactt	gcattgattt	gacttttaat	9420
tggtgagcacc	gtattctact	aggaagagca	agagagggtg	tgtggatcct	tgtgtccatg	9480
gaccgaccat	gggaggcttc	gctatccatc	catattttta	cggaagtatt	aaaaggaatt	9540
ctaactagat	ccaaaagatt	catttttact	ttgatggcag	tgattatggg	cctcattgca	9600
gtcacagcta	ctgctgcggc	tgctggaatt	gctttacact	cctctgttca	aactgcagaa	9660
tacgtaaaatg	aattggcaaaa	gaattcctca	aaattgtgga	attctcagat	ccaaatagat	9720
caaaaattgg	caaaacaaat	taatgatctt	agacaaactg	tcatttggtat	gggagaggct	9780
catgagcttg	gaatatcttt	ttcagttacg	atgtgactgg	aatacatcag	atttttgtgt	9840
tacaccacaa	gcctataatg	agtctgagca	tcactgggac	atggttagat	gccatctgca	9900
aggaggagaa	gataatctta	ctttagacat	ttcaaaaatta	aaagaatttt	ttttttcttt	9960
gagacagagt	ctcgctctgt	cgcccaggct	ggagtgcagt	ggcgtgatct	cagctcactg	10020
caagttccgc	ctcctggggt	tacaccattc	tcctgcctga	gcctcccaag	tagttgggag	10080
tacaggagcc	caccaccatg	cctggctaatt	tttttttggg	tttttaatat	agatggagtt	10140
tcaccgtggt	agccaggatg	gtctcgatct	cctgaccttg	tgatctgccc	accttgccct	10200
cccaaagtgc	tgggattaca	gtcgtgagcc	accgtgcca	gccaagaaaa	aatttttgag	10260
gcatcaaaaag	cccatttaaa	tttggtgcca	ggaacggaga	caatcgtgaa	agctgctgat	10320
agcttcacaa	atcttaagcc	agtcacttgg	gttaaaaagca	tcagaagtgt	cactattgta	10380
aatttcatat	taatccttgt	atgcctgttc	tgtctgttgt	tagtctacag	gtgtatccag	10440
cagctccaaa	gagacagcaa	ccagcaagaa	tgggccatag	tgacgatggt	ggttttgtca	10500
aaaagaaaag	ggggggatat	gtaaggaaaa	gagagatcag	actttcactg	tgtctatgta	10560
gaaaagggaag	acataagaaa	ctccattttg	atctgtacta	agaaaaattg	ttttgccttg	10620
agatgctgtt	aatctgtaac	tttagcccca	accctgtgct	cacggaaaca	tgtgctgtaa	10680
ggtttaaggg	atctagggct	gtgcaggatg	taccttggtta	acaatatggt	tgcaggcagt	10740
atgttttgta	aaagtcactg	ccattctcca	ttctcgatta	accaggggct	caatgcactg	10800
tggaaaagcca	caggaaacctc	tgcccaagaa	agcctggctg	ttgtgggaag	tcagggaccc	10860
cgaatggagg	gaccagctgg	tgctgcatca	ggaaacataa	attgtgaaga	tttcttggtg	10920
atttatcagt	ttccaaaatt	aatactttta	taatttctta	cacctgtctt	actttaatct	10980
cttaatcctg	ttatctttgt	aagctgagga	tatacgtcac	ctcaggacca	ctattgtaca	11040
aattgattgt	aaaacatggt	cacatgtggt	tgaacaatat	gaaatcagtg	caccttgaaa	11100
atgaacagaa	taacagtgat	tttagggaac	aaaggaagac	aaccataagg	tctgactgcc	11160

Substitute Sequence Listing_USSN 10587032_PP019482.007

tgaggggtcg	ggcaaaaagc	catatttttc	ttcttgcaga	gagcctataa	atggacgtgc	11220
aagtaggaga	gatattgcta	aattcttttc	ctagcaagga	atataatact	aagaccctag	11280
ggaaagaatt	gatttcctgg	ggggaggtct	ataaacggcc	gctctgggag	tgtctgtcct	11340
atgtggttga	gataaggact	gagatacgcc	ctggtctcct	gcagtaccct	caggcttact	11400
aggattggga	aaccccagtc	ctggtaaatt	tgaggtcagg	ccggttcttt	gctctgaacc	11460
ctgttttctg	ttaagatggt	tatcaagaca	atacatgcac	cgctgaacat	agacccttat	11520
caggagtttc	tgattttgct	ctggtcctgt	ttcttcagaa	gcattgtcatc	tttgcctctgc	11580
cttctgcctc	ttgaagcatg	tgatctttgt	gacctactcc	ctgttcatac	acccctcccc	11640
ttttaaaatc	cctaataaaa	acttgctggg	tttgtggctc	aggggggcat	catggaccta	11700
ccaatacgtg	atgtcacccc	cgggtggccc	gctgtaaaat	tcctttcttt	atactcttat	11760
ttctcagacc	agctgacact	tagggaaaat	agaaagaacc	tatgttgaaa	tattggaggc	11820
gggttccccc	gatacctggg	tattgtccaa	ggtttccttt	gctgaggagg	attagtaaaa	11880
ggaatgcctc	catctcctgc	atgtccctgg	gaacagaatg	ttcccaccaa	ccaccctgtg	11940
gctggaggcg	ggaatgctg	gcagcaatgc	tgctctatta	ctctttgcta	cactgagatg	12000
tttgggtgga	gagaagcata	aatctggcct	ctgggcacag	caccttcctt	12060	
tgaacttatt	tgtgacacag	attcctttgc	tcacgttttc	ctgttgactt	tctcaccact	12120
caccctattc	tcctgtggca	ttcgccctgc	ggagatagtg	aaaatagtaa	taaatactga	12180
gggaactcag	actgagggaa	ctcagactgg	gcagaccggg	gccagtgtgg	gtcctccata	12240
tgctgagcgc	cggttccctg	ggcccactgt	tctttctcta	tactttgtct	ctgtgcctta	12300
ttttctcagt	ctctcattcc	acctgatgag	aaatacccac	aggtgtggag	gggctggccc	12360
ccttca						12366

<210> 76

<211> 2148

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 76

atggggcaaa	ctgaaagtaa	atatgcctct	tatctcagct	ttattaaaaat	tcttttaaga	60
agagggggag	ttagagcttc	tacagaaaat	ctaattacgc	tatttcaaac	aatagaacaa	120
ttctgcccac	ggtttccaga	acagggaaact	ttagatctaa	aagattggga	aaaaattggc	180
aaagaattaa	aacaagcaaa	tagggaaggt	aaaatcatcc	cacttacagt	atggaatgat	240
tgggccatta	ttaaagcaac	tttagaacca	tttcaaacag	gagaagatat	tgtttcagtt	300
tctgatgccc	ctaaaagctg	tgtaacagat	tgtgaagaag	aggcagggac	agaatcccag	360
caaggaacgg	aaagtccaca	ttgtaaatat	gtagcagagt	ctgtaatggc	tcagtcaacg	420
caaaatgttg	actacagtca	attacaggag	ataatatacc	ctgaatcatc	aaaattgggg	480
gaaggagggtc	cagaatcatt	ggggccatca	gagcctaaac	cacgatcgcc	atcaactcct	540
cctcccgtgg	ttcagatgcc	tgtaacatta	caacctcaaa	cgcaggttag	acaagcacaa	600
accccaagag	aaaatcaagt	agaaaggga	agagtctcta	tcccggcaat	gccaactcag	660
atacagtatc	cacaatatca	gccggtagaa	aataagacc	aaccgctggt	agtttatcaa	720
taccggctgc	caaccgagct	tcagtatcgg	cctccttcag	aggttcaata	cagacctcaa	780
gcggtgtgtc	ctgtgccaaa	tagcacggca	ccataccagc	aaccacagc	gatggcgtct	840
aattcaccag	caacacagga	cgcggcgctg	tatcctcagc	cgcccactgt	gagacttaat	900
cctacagcat	cacgtagtgg	acaggggtgt	gcactgcattg	cagtcattga	tgaagccaga	960
aaacaggggc	atcttgaggc	atggcggttc	ctggttaatt	tacaactggt	acaggccggg	1020
gaagagactc	aagttaggag	gcctgcccga	gctgagacta	gatgtgaacc	tttcaccatg	1080
aaaatgttaa	aagatataaa	ggaaggagtt	aaacaatatg	gatccaactc	cccttatata	1140
agaacattat	tagattccat	tgctcatgga	aatagactta	ctccttatga	ctgggaaatt	1200
ttggccaaat	cttccctttc	atcctctcag	tatctacagt	ttaaaacctg	gtggattgat	1260
ggagtacaa	aacaggtacg	aaaaaatcag	gctactaagc	ccactgttaa	tatagacgca	1320
gaccaattgt	taggaacagg	tccaaattgg	agcaccatta	accaacaatc	agtgatgcag	1380
aatgaggcta	ttgaacaagt	aagggtatatt	tgccctcagg	cctggggaaa	aattcaggac	1440
ccaggaacag	ctttccctat	taattcaatt	agacaaggct	ctaaagagcc	atatcctgac	1500
tttgtggcaa	gattacaaga	tgctgtctaa	aagtctatta	cagatgacaa	tgcccgaaaa	1560
gttattgtag	aattaatggc	ctatgaaaat	gcaaatccag	aatgtcagtc	ggccataaag	1620
ccattaaaag	gaaaagtccc	agcaggagtt	gatgtaatga	cagaatatgt	gaaggcttgt	1680
gatgggattg	gaggagctat	gcataaggca	atgctaattg	ctcaagcaat	gagggggctc	1740
actctaggag	gacaagttag	aacatttggg	aaaaaatgtt	ataattgtgg	tcaaatcggt	1800
catctgaaaa	ggagtgtccc	agtcttaaat	aaacagaata	taataaatca	agctattaca	1860
gcaaaaaata	aaaagccatc	tggcctgtgt	ccaaaatgtg	gaaaaggaaa	acattggggc	1920
aatcaatgtc	attctaaatt	tgataaggat	gggcaaccat	tgctgggaaa	caggaagagg	1980
ggccagcctc	aggcccccca	acaaactggg	gcattcccag	ttcaactgtt	tgttcctcag	2040
ggttttcaag	gacaacaacc	cctacagaaa	ataccaccac	ttcaggaggat	cagccaatta	2100

Substitute Sequence Listing_USSN 10587032_PP019482.007

caacaatcca acagctgtcc cgcgccacag caggcagcac cgcagtaa 2148

<210> 77
<211> 2151
<212> DNA
<213> Artificial Sequence

<220>
<223> Manipulated Gag

<400> 77
atgggccaga ccgagagcaa gtacgccagc tacctgagct tcatcaagat cctgctgcgc 60
cgcgcgccg tgccgcccag caccgagaac ctgataccac tgttccagac catcgagcag 120
ttctgcccc gggtccccga gcagggcacc ctggacctga aggactggga gaagatcgcc 180
aaggagctga agcaggccaa ccgcgagggc aagatcatcc ccctgaccgt gtggaacgac 240
tgggccatca tcaaggccac cctggagccc ttccagaccg gcgaggacat cgtgagcgtg 300
agcgacgccc ccaagagctg cgtgaccgac tgcgaggagg aggccggcac cgagagccag 360
cagggcaccg agagcagcca ctgcaagtac gtggccgaga gcgtgatggc ccagagcacc 420
cagaacgtgg actacagcca gctgcaggag atcatctacc ccgagagcag caagctgggc 480
gaggcgccg ccgagagcct ggccccagc gagcccaagc cccgcagccc cagcaccccc 540
cccccgctgg tgcagatgcc cgtgaccctg cagccccaga cccaggtgcg ccaggcccag 600
accccccgcg agaaccaggt ggagcgcgac cgcgtgagca tccccgcat gccaccccag 660
atccagtacc cccagtacca gcccgctggag aacaagacc agcccctggt ggtgtaccag 720
taccgcctgc ccaccgagct gcagtaccgc cccccagcg aggtgcagta ccgccccag 780
gccgtgtgcc ccgtgcccac cagcaccgcc cctaccagc agcccaccgc catggccagc 840
aacagcccc ccaccagga cgcgcctctg tacccccag ccccaccgt gcgcctgaac 900
cccaccgcca gccgcagcgg ccaggcgccg gccctgcacg ccgtgatcga cgaggcccg 960
aagcaggcg acctggaggc ctggcgcttc ctggtgatcc tgcagctggt gcaggccggc 1020
gaggagacc aggtgggcgc cccgcccgc gccgagacc gctgcgagcc cttcaccatg 1080
aagatgctga aggacatcaa ggagggcgtg aagcagtagc gcagcaacag cccctacatc 1140
cgaccctgc tggacagcat cgcccacggc aaccgcctga cccctacga ctgggagatc 1200
ctggccaaga gcagcctgag cagcagccag tacctgcagt tcaagacctg gtggatcgac 1260
ggcgtgcagg agcaggtgcg caagaaccag gccaccaagc ccaccgtgaa catcgacgcc 1320
gaccagctgc tgggcaccgg cccaacttg agcaccatca accagcagag cgtgatgcag 1380
aacgaggcca tcgagcaggt gcgcgccatc tgcctgcgcg cctggggcaa gatccaggac 1440
cccgccaccg ccttccccat caacagcatc cgccagggca gcaaggagcc ctaccccagc 1500
ttcgtggccc gcctgcagga cgccgcccag aagagcatca ccgacgaca cgcccgaag 1560
gtgatcgtg agctgatggc ctacgagaac gccaacccc agtgccagag cgccatcaag 1620
cccctgaagg gcaagggtgc cgccggcgtg tacctgatca ccgagtacgt gaaggcctgc 1680
gacggcatcg gcggcgccat gcacaaggcc atgctgatgg ccaggccat gcgcggcctg 1740
accctgggcg gccaggtgcg caccttcggc aagaagtgt acaactgcg ccagatcgcc 1800
cactgaagc gcagctgccc cgtgctgaac aagcagaaca tcatcaacca ggccatcacc 1860
gccaagaaca agaagcccag cggcctgtgc cccaagtgc gcaagggcaa gcaactggcc 1920
aaccagtgc acagcaagtt cgacaaggac ggccagcccc tgagcggcaa ccgcaagcgc 1980
ggccagcccc agggccccca gcagaccggc gccttcccc tgcagctgtt cgtgccccag 2040
ggcttccagg gccagcagcc cctgcagaag atccccccc tgcaggcggt gagccagctg 2100
cagcagagca acagctgccc cgccccccag caggccgccc cccaggctta a 2151

<210> 78
<211> 715
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 78
Met Gly Gln Thr Glu Ser Lys Tyr Ala Ser Tyr Leu Ser Phe Ile Lys
1 5 10 15
Ile Leu Leu Arg Arg Gly Gly Val Arg Ala Ser Thr Glu Asn Leu Ile
20 25 30
Thr Leu Phe Gln Thr Ile Glu Gln Phe Cys Pro Trp Phe Pro Glu Gln
35 40 45

Substitute Sequence Listing_USSN 10587032_PP019482.007

Gly Thr Leu Asp Leu Lys Asp Trp Glu Lys Ile Gly Lys Glu Leu Lys
50 55 60

Gln Ala Asn Arg Glu Gly Lys Ile Ile Pro Leu Thr Val Trp Asn Asp
65 70 75 80

Trp Ala Ile Ile Lys Ala Thr Leu Glu Pro Phe Gln Thr Gly Glu Asp
85 90 95

Ile Val Ser Val Ser Asp Ala Pro Lys Ser Cys Val Thr Asp Cys Glu
100 105 110

Glu Glu Ala Gly Thr Glu Ser Gln Gln Gly Thr Glu Ser Ser His Cys
115 120 125

Lys Tyr Val Ala Glu Ser Val Met Ala Gln Ser Thr Gln Asn Val Asp
130 135 140

Tyr Ser Gln Leu Gln Glu Ile Ile Tyr Pro Glu Ser Ser Lys Leu Gly
145 150 155 160

Glu Gly Gly Pro Glu Ser Leu Gly Pro Ser Glu Pro Lys Pro Arg Ser
165 170 175

Pro Ser Thr Pro Pro Pro Val Val Gln Met Pro Val Thr Leu Gln Pro
180 185 190

Gln Thr Gln Val Arg Gln Ala Gln Thr Pro Arg Glu Asn Gln Val Glu
195 200 205

Arg Asp Arg Val Ser Ile Pro Ala Met Pro Thr Gln Ile Gln Tyr Pro
210 215 220

Gln Tyr Gln Pro Val Glu Asn Lys Thr Gln Pro Leu Val Val Tyr Gln
225 230 235 240

Tyr Arg Leu Pro Thr Glu Leu Gln Tyr Arg Pro Pro Ser Glu Val Gln
245 250 255

Tyr Arg Pro Gln Ala Val Cys Pro Val Pro Asn Ser Thr Ala Pro Tyr
260 265 270

Gln Gln Pro Thr Ala Met Ala Ser Asn Ser Pro Ala Thr Gln Asp Ala
275 280 285

Ala Leu Tyr Pro Gln Pro Pro Thr Val Arg Leu Asn Pro Thr Ala Ser
290 295 300

Arg Ser Gly Gln Gly Gly Ala Leu His Ala Val Ile Asp Glu Ala Arg
305 310 315 320

Lys Gln Gly Asp Leu Glu Ala Trp Arg Phe Leu Val Ile Leu Gln Leu
325 330 335

Val Gln Ala Gly Glu Glu Thr Gln Val Gly Ala Pro Ala Arg Ala Glu
340 345 350

Thr Arg Cys Glu Pro Phe Thr Met Lys Met Leu Lys Asp Ile Lys Glu
355 360 365

Gly Val Lys Gln Tyr Gly Ser Asn Ser Pro Tyr Ile Arg Thr Leu Leu
370 375 380

Substitute Sequence Listing_USSN 10587032_PP019482.007

Asp Ser Ile Ala His Gly Asn Arg Leu Thr Pro Tyr Asp Trp Glu Ile
 385 390 395 400
 Leu Ala Lys Ser Ser Leu Ser Ser Ser Gln Tyr Leu Gln Phe Lys Thr
 405 410 415
 Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Lys Asn Gln Ala Thr
 420 425 430
 Lys Pro Thr Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Thr Gly Pro
 435 440 445
 Asn Trp Ser Thr Ile Asn Gln Gln Ser Val Met Gln Asn Glu Ala Ile
 450 455 460
 Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Gly Lys Ile Gln Asp
 465 470 475 480
 Pro Gly Thr Ala Phe Pro Ile Asn Ser Ile Arg Gln Gly Ser Lys Glu
 485 490 495
 Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Ala Ala Gln Lys Ser
 500 505 510
 Ile Thr Asp Asp Asn Ala Arg Lys Val Ile Val Glu Leu Met Ala Tyr
 515 520 525
 Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys Gly
 530 535 540
 Lys Val Pro Ala Gly Val Asp Val Ile Thr Glu Tyr Val Lys Ala Cys
 545 550 555 560
 Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln Ala
 565 570 575
 Met Arg Gly Leu Thr Leu Gly Gly Gln Val Arg Thr Phe Gly Lys Lys
 580 585 590
 Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Arg Ser Cys Pro Val
 595 600 605
 Leu Asn Lys Gln Asn Ile Ile Asn Gln Ala Ile Thr Ala Lys Asn Lys
 610 615 620
 Lys Pro Ser Gly Leu Cys Pro Lys Cys Gly Lys Gly Lys His Trp Ala
 625 630 635 640
 Asn Gln Cys His Ser Lys Phe Asp Lys Asp Gly Gln Pro Leu Ser Gly
 645 650 655
 Asn Arg Lys Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
 660 665 670
 Pro Val Gln Leu Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Leu
 675 680 685
 Gln Lys Ile Pro Pro Leu Gln Gly Val Ser Gln Leu Gln Gln Ser Asn
 690 695 700
 Ser Cys Pro Ala Pro Gln Gln Ala Ala Pro Gln
 705 710 715

Substitute Sequence Listing_USSN 10587032_PP019482.007

<210> 79
 <211> 716
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Manipulated Gag

<400> 79

```

Met Gly Gln Thr Glu Ser Lys Tyr Ala Ser Tyr Leu Ser Phe Ile Lys
1      5      10      15
Ile Leu Leu Arg Arg Gly Gly Val Arg Ala Ser Thr Glu Asn Leu Ile
20     25     30
Thr Leu Phe Gln Thr Ile Glu Gln Phe Cys Pro Trp Phe Pro Glu Gln
35     40     45
Gly Thr Leu Asp Leu Lys Asp Trp Glu Lys Ile Gly Lys Glu Leu Lys
50     55     60
Gln Ala Asn Arg Glu Gly Lys Ile Ile Pro Leu Thr Val Trp Asn Asp
65     70     75     80
Trp Ala Ile Ile Lys Ala Thr Leu Glu Pro Phe Gln Thr Gly Glu Asp
85     90     95
Ile Val Ser Val Ser Asp Ala Pro Lys Ser Cys Val Thr Asp Cys Glu
100    105    110
Glu Glu Ala Gly Thr Glu Ser Gln Gln Gly Thr Glu Ser Ser His Cys
115    120    125
Lys Tyr Val Ala Glu Ser Val Met Ala Gln Ser Thr Gln Asn Val Asp
130    135    140
Tyr Ser Gln Leu Gln Glu Ile Ile Tyr Pro Glu Ser Ser Lys Leu Gly
145    150    155    160
Glu Gly Gly Pro Glu Ser Leu Gly Pro Ser Glu Pro Lys Pro Arg Ser
165    170    175
Pro Ser Thr Pro Pro Pro Val Val Gln Met Pro Val Thr Leu Gln Pro
180    185    190
Gln Thr Gln Val Arg Gln Ala Gln Thr Pro Arg Glu Asn Gln Val Glu
195    200    205
Arg Asp Arg Val Ser Ile Pro Ala Met Pro Thr Gln Ile Gln Tyr Pro
210    215    220
Gln Tyr Gln Pro Val Glu Asn Lys Thr Gln Pro Leu Val Val Tyr Gln
225    230    235    240
Tyr Arg Leu Pro Thr Glu Leu Gln Tyr Arg Pro Pro Ser Glu Val Gln
245    250    255
Tyr Arg Pro Gln Ala Val Cys Pro Val Pro Asn Ser Thr Ala Pro Tyr
260    265    270
Gln Gln Pro Thr Ala Met Ala Ser Asn Ser Pro Ala Thr Gln Asp Ala
275    280    285
    
```

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ala Leu Tyr Pro Gln Pro Pro Thr Val Arg Leu Asn Pro Thr Ala Ser
290 295 300

Arg Ser Gly Gln Gly Gly Ala Leu His Ala Val Ile Asp Glu Ala Arg
305 310 315 320

Lys Gln Gly Asp Leu Glu Ala Trp Arg Phe Leu Val Ile Leu Gln Leu
325 330 335

Val Gln Ala Gly Glu Glu Thr Gln Val Gly Ala Pro Ala Arg Ala Glu
340 345 350

Thr Arg Cys Glu Pro Phe Thr Met Lys Met Leu Lys Asp Ile Lys Glu
355 360 365

Gly Val Lys Gln Tyr Gly Ser Asn Ser Pro Tyr Ile Arg Thr Leu Leu
370 375 380

Asp Ser Ile Ala His Gly Asn Arg Leu Thr Pro Tyr Asp Trp Glu Ile
385 390 395 400

Leu Ala Lys Ser Ser Leu Ser Ser Ser Gln Tyr Leu Gln Phe Lys Thr
405 410 415

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Lys Asn Gln Ala Thr
420 425 430

Lys Pro Thr Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Thr Gly Pro
435 440 445

Asn Trp Ser Thr Ile Asn Gln Gln Ser Val Met Gln Asn Glu Ala Ile
450 455 460

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Gly Lys Ile Gln Asp
465 470 475 480

Pro Gly Thr Ala Phe Pro Ile Asn Ser Ile Arg Gln Gly Ser Lys Glu
485 490 495

Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Ala Ala Gln Lys Ser
500 505 510

Ile Thr Asp Asp Asn Ala Arg Lys Val Ile Val Glu Leu Met Ala Tyr
515 520 525

Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys Gly
530 535 540

Lys Val Pro Ala Gly Val Asp Val Ile Thr Glu Tyr Val Lys Ala Cys
545 550 555 560

Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln Ala
565 570 575

Met Arg Gly Leu Thr Leu Gly Gly Gln Val Arg Thr Phe Gly Lys Lys
580 585 590

Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Arg Ser Cys Pro Val
595 600 605

Leu Asn Lys Gln Asn Ile Ile Asn Gln Ala Ile Thr Ala Lys Asn Lys
610 615 620

Substitute Sequence Listing_USSN 10587032_PP019482.007

Lys Pro Ser Gly Leu Cys Pro Lys Cys Gly Lys Gly Lys His Trp Ala
625 630 635 640

Asn Gln Cys His Ser Lys Phe Asp Lys Asp Gly Gln Pro Leu Ser Gly
645 650 655

Asn Arg Lys Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
660 665 670

Pro Val Gln Leu Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Leu
675 680 685

Gln Lys Ile Pro Pro Leu Gln Gly Val Ser Gln Leu Gln Gln Ser Asn
690 695 700

Ser Cys Pro Ala Pro Gln Gln Ala Ala Pro Gln Ala
705 710 715

<210> 80
<211> 6486
<212> DNA
<213> Artificial Sequence

<220>
<223> pCMVKm2.gagopt PCAV vector

<400> 80
gccgcggaat ttcgactcta ggccattgca tacgttgtat ctatatcata atatgtacat 60
ttatattggc tcatgtccaa tatgaccgcc atgttgacat tgattattga ctagtattata 120
atagtaatca attacggggt cattagttca tagcccatat atggagttcc gcgttacata 180
acttacgta aatggccgc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat 240
aatgacgat gttcccatg taacgccaat agggactttc cattgacgtc aatgggtgga 300
gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc caagtccgcc 360
ccctattgac gtcaatgacg gtaaatggcc cgctggcat tatgccagt acatgacctt 420
acgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta ccatggtgat 480
gcggttttgg cagtacacca atgggcgtgg atagcggttt gactcacggg gatttccaag 540
tctccacccc attgacgtca atgggagttt gttttggcac caaatcaac gggactttcc 600
aaaatgtcgt aataaccccg ccccgttgac gcaaattggc ggtaggcgtg tacggtggga 660
ggtctatata agcagagctc gtttagtgaa ccgtcagatc gcctggagac gccatccacg 720
ctgttttgac ctccatagaa gacaccggga ccgatccagc ctccgcggcc gggaacggtg 780
cattggaacg cggtattccc gtgccaagag tgacgtaagt accgcctata gactctatag 840
gcacacccct ttggctctta tgcattgctat actgtttttg gcttggggcc tatacacc 900
cgcttcctta tgctataggt gatggtatag cttagcctat aggtgtgggt tattgacct 960
tattgaccac tcccctattg gtgacgatac ttccattac taatccataa catggctctt 1020
tgccacaact atctctattg gctatatgcc actactctgt ccttcagaga ctgacacgga 1080
ctctgtattt ttacaggatg ggttcccatt tattatttac aaattcacat atacaacaac 1140
gccgtccccc gtgcccgcag tttttattaa acatagcgtg ggatctccac gcgaatctcg 1200
ggtacgtgtt ccggacatgg gctcttctcc ggtagcggcg gagcttccac atccgagccc 1260
tggtcccatg cctccagcgg ctcatggctc ctcggcagct ccttgctcct aacagtggag 1320
gccagactta ggcacagcac aatgcccacc accaccagtg tgccgcacaa ggccgtggcg 1380
gtagggtatg tgtctgaaaa tgagctcgga gattgggtc gcaccgctga cgcagatgga 1440
agacttaagg cagcggcaga agaagatgca ggcagctgag ttgttgtatt ctgataagag 1500
tcagaggtaa ctcccgttg ggtgctgtta acggtggagg gcagtgtagt ctgagcagta 1560
ctcgttgctg ccgcgcgcgc caccagacat aatagctgac agactaacag actgttcctt 1620
tccatgggtc ttttctgcag tcaccgtcgt cgacgccacc atgggcccaga ccgagagcaa 1680
gtacgccagc tacctgagct tcatcaagat cctgctgcgc cgcgccggcg tgcgcgccag 1740
caccagaac ctgatcacc tgttcagac catcgacagc ttctgcccct ggttccccga 1800
gcagggcacc ctggacctga aggactggga gaagatcggc aaggagctga agcaggccaa 1860
ccgcgagggc aagatcatcc ccctgaccgt gtggaacgac tgggcatca tcaaggccac 1920
cctggagccc ttccagaccg gcgaggacat cgtgagcgtg agcgacgcc ccaagagctg 1980
cgtgaccgac tgcgaggagg aggccggcac cgagagccag cagggcaccg agagcagcca 2040
ctgcaagtac gtggccgaga gcgtgatggc ccagagcacc cagaacgtgg actacagcca 2100
gctgcaggag atcatctacc ccgagagcag caagctgggc gagggcggcc ccgagagcct 2160

Substitute Sequence Listing_USSN 10587032_PP019482.007

gggccccagc	gagcccaagc	cccgcagccc	cagcaccccc	cccccggtg	tgcagatgcc	2220
cgtgaccctg	cagccccaga	cccaggtgcg	ccaggcccag	accccccgcg	agaaccaggt	2280
ggagcgcgac	cgcgtagca	tccccgccat	gcccaccag	atccagtacc	cccagtagca	2340
gcccgtggag	aacaagaccc	agcccctggt	ggtgtaccag	taccgcctgc	ccaccgagct	2400
gcagtaccgc	ccccccagcg	aggtgcagta	ccgccccag	gccgtgtgcc	ccgtgccccaa	2460
cagcaccgcc	ccctaccagc	agcccaccgc	catggccagc	aacagccccg	ccaccagga	2520
cgccgcccgt	tacccccagc	ccccaccgt	gcgcctgaac	cccaccgcca	gccgcagcgg	2580
ccagggcggc	gccctgcacg	ccgtgatcga	cgaggcccgc	aagcagggcg	acctggaggc	2640
ctggcgcttc	ctggtgatcc	tgcagctggt	gcaggccggc	gaggagacc	aggtgggcgc	2700
ccccgcccgc	gccgagaccc	gctgcgagcc	cttcaccatg	aagatgctga	aggacatcaa	2760
ggagggcggt	aagcagtagc	gcagcaacag	cccctacatc	cgcaccctgc	tggacagcat	2820
cgcccacggc	aaccgcctga	ccccctacga	ctgggagatc	ctggccaaga	gcagcctgag	2880
cagcagccag	tacctgcagt	tcaagacctg	gtggatcgac	ggcgtgcagg	agcaggtgcg	2940
caagaaccag	gccaccaagc	ccaccgtgaa	catcgacgcc	gaccagctgc	tgggcaccgg	3000
ccccaaactgg	agcaccatca	accagcagag	cgtgatgcag	aacgaggcca	tcgagcaggt	3060
gcgcgccatc	tgcctgcgcg	cctggggcaa	gatccaggac	cccggcaccg	ccttccccat	3120
caacagcatc	cgccagggca	gcaaggagcc	ctaccccagc	ttcgtggccc	gcctgcagga	3180
cgccgcccag	aagagcatca	ccgacgacaa	cgcccgaag	gtgatcgtgg	agctgatggc	3240
ctacgagaac	gccaaccccc	agtgccagag	cgccatcaag	cccctgaagg	gcaaggtgcc	3300
cgccggcggt	gacgtgatca	ccgagtacgt	gagggccatg	gacggcatcg	gcggcgccat	3360
gcacaaggcc	atgctgatgg	cccaggccat	gcgcggcctg	accctggggc	gccaggtgcg	3420
caccttcggc	aagaagtgtc	acaactgcgg	ccagatcggc	cacctgaagc	gcagctgccc	3480
cgtgctgaac	aagcagaaca	tcatcaacca	ggccatcacc	gccaagaaca	agaagcccag	3540
cggcctgtgc	cccaagtgcg	gcaagggcaa	gcactggggc	aaccagtgcc	acagcaagtt	3600
cgacaaggac	ggccagcccc	tgagcggcaa	ccgcaagcgc	ggccagcccc	aggcccccca	3660
gcagaccggc	gccttccccg	tgcagctggt	gagggccatg	ggcttccagg	gccagcagcc	3720
cctgcagaag	atcccccccc	tgcaggcggt	gagccagctg	cagcagagca	acagctgccc	3780
cgccccccag	caggccgccc	cccaggctta	agaattcaga	ctcagcaag	tctagaaagc	3840
catggatatt	ggatccacta	cgcgttagag	ctcgtgatc	agcctcgact	gtgccttcta	3900
gttgccagcc	atctgttgtt	tgccccctcc	ccgtgccttc	cttgaccctg	gaaggtgcca	3960
ctcccactgt	cctttcctaa	taaaatgagg	aaattgcatt	gcattgtctg	agtaggtgtc	4020
attctattct	ggggggtagg	gtggggcagg	acagcaagg	ggaggattgg	gaagacaata	4080
gcaggggggt	gggcgaagaa	gtccagcatg	agatccccgc	gctggaggat	catccagccg	4140
gcgtcccgga	aaacgattcc	gaagcccaac	ctttcataga	aggcggcggt	ggaatcgaaa	4200
tctcgtgatg	gcaggttggg	cgtcgcttgg	tcggtcattt	cgaaccccag	agtcccgcct	4260
agaagaactc	gtcaagaagg	cgatagaagg	cgatgcgctg	cgaatcggga	gcggcgatac	4320
cgtaaaacac	gaggaagcgg	tcagcccatt	cgccgccaag	ctcttcagca	atatcacggg	4380
tagccaaacgc	tatgtcctga	tagcggtccg	ccacaccag	ccggccacag	tcgatgaatc	4440
cagaaaagcg	gccattttcc	accatgatat	tcggcaagca	ggcatcgcca	tgggtcacga	4500
cgagatcctc	gccgtcgggc	atgcgcgcct	tgagcctggc	gaacagttcg	gctggcgcca	4560
gcccctgatg	ctcttcgtcc	agatcatcct	gatcgacaag	accggcttcc	atccgagtag	4620
gtgctcgctc	gatgcgatgt	ttcgcttggt	ggtcgaatgg	gcaggtagcc	ggatcaagcg	4680
tatgcagccg	ccgcattgca	tcagccatga	tggatacttt	ctcggcagga	gcaaggtgag	4740
atgacaggag	atcctgcccc	ggcacttcgc	ccaatagcag	ccagtccctt	cccgttccag	4800
tgacaacgct	gagcacagct	gcgcaaggaa	cggcgctcgt	ggccagccac	gatagccgcg	4860
ctgcctcgct	ctgcagttca	ttcagggcac	cggacaggtc	ggtcttgaca	aaaagaaccg	4920
ggcgcccctg	cgctgacagc	cggaacacgg	cggcatcaga	gcagccgatt	gtctgttgtg	4980
cccagtcata	gccgaatagc	ctctccaccc	aagcggccgg	agaacctgcg	tgcaatccat	5040
cttgttcaat	catgcgaaac	gatcctcatc	ctgtctcttg	atcagatctt	gatccccctg	5100
gccatcagat	ccttggcggc	aagaaagcca	tccagtttac	tttgaggggc	ttcccaacct	5160
taccagaggg	cgccccagct	ggcaattccg	gttcgcttgc	tgtccataaa	accgcccagt	5220
ctagctatcg	ccatgtaagc	ccactgcaag	ctacctgctt	tctctttgcg	cttgcgtttt	5280
cccttgcca	gatagcccag	tagctgacat	tcacccgggg	tcagcaccgt	ttctgcggac	5340
tggctttcta	cgtgttccgc	ttcctttagc	agcccttgcg	ccctgagtag	ttgcggcagc	5400
gtgaagctaa	ttcatggtta	aatttttgtt	aatcagctc	attttttaac	caataggccg	5460
aaatcggcaa	aatcccttat	aatcaaaaag	aatagcccca	gatagggttg	agtgttgttc	5520
cagttttgaa	caagagtcca	ctattaaaga	acgtgacatc	caacgtcaaa	gggcgaaaaa	5580
ccgtctatca	gggcgatggc	cggatcagct	tatgcgggtg	gaaataccgc	acagatgcgt	5640
aaggagaaaa	taccgcatca	ggcgctcttc	cgcttctctg	ctcactgact	cgctgcgctc	5700
ggtcggttcg	ctgcggcgag	cggatcagc	tcactcaaa	gcggtaatac	ggttatccac	5760
agaatcaggg	gataacgcag	gaaagaacat	gtgagcaaaa	ggccagcaaa	aggccaggaa	5820
ccgtaaaaaa	gccgcgttgc	tggcgttttt	ccataggtct	cgccccctg	acgagcatca	5880
caaaaatcga	cgctcaagtc	agaggtggcg	aaacccgaca	ggactataaa	gataccaggc	5940

Substitute Sequence Listing_USSN 10587032_PP019482.007

gtttccccct	ggaagctccc	tcgtgcgctc	tcctgtttccg	accctgccgc	ttaccggata	6000
cctgtccgcc	tttctccctt	cggaagcgt	ggcgctttct	catagctcac	gctgtaggta	6060
tctcagttcg	gtgtaggctg	ttcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	6120
gcccgaaccg	tgcgccttat	ccgtaacta	tcgtcttgag	tccaacccgg	taagacacga	6180
cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	agagcgaggt	atgtaggcgg	6240
tgctacagag	ttcttgaagt	ggtggcctaa	ctacggctac	actagaagga	cagtatttgg	6300
tatctgcgct	ctgctgaagc	cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	6360
caaacaaacc	accgctggta	gcggtggttt	ttttgtttgc	aagcagcaga	ttacgcgcag	6420
aaaaaaagga	tctcaagaag	atcctttgat	cttttctact	gaacggtgat	ccccaccgga	6480
attgcg						6486

<210> 81
 <211> 2103
 <212> DNA
 <213> Human endogenous retrovirus, K family (HERV-K)

<400> 81						
atgaacccaa	gcgagatgca	aagaaaagca	cctccgcgga	gacggagaca	tcgcaatcga	60
gcaccgttga	ctcacaagat	gaacaaaatg	gtgacgtcag	agaacagat	gaagttgcc	120
tccaccaaga	aggcagagcc	gccaaacttg	gcacaactaa	agaagctgac	gcagttagct	180
acaaaatatt	tagagaacac	aaaggtgaca	caaaccctag	agagtatgct	gcttgcagcc	240
ttgatgattg	tatcaatggt	ggtaagtctc	cctatgcctg	caggagcagc	tgcagctaac	300
tatacctact	gggcctatgt	gcctttcccg	cccttaattc	gggcagtcac	atggatggat	360
aatcctacag	aagtatatgt	taatgatagt	gtatgggtac	ctggccccc	agatgatcgc	420
tgccctgcc	aacctgagga	agaagggatg	atgataaata	tttccattgg	gtatcattat	480
cctcctattt	gcctagggag	agcaccagga	tgtttaatgc	ctgcagtcca	aaattggtg	540
gtagaagtac	ctactgtcag	tcccatctgt	agattcactt	atcacatggt	aagcgggatg	600
tcactcaggc	cacgggtaaa	ttattttaca	gacttttctt	atcaaagatc	attaaaaatt	660
agacctaaag	ggaaaccttg	ccccaaggaa	attcccaaag	aatcaaaaaa	tacagaagtt	720
ttagtttggg	aagaatgtgt	ggccaatagt	gcggtgat	tacaaaacaa	tgaattcgga	780
actattatag	attgggcacc	tcgaggtcaa	ttctaccaca	attgctcagg	acaaactcag	840
tcgtgtccaa	gtgcacaagt	gagtcacagt	gttgatagcg	acttaacaga	aagtttagac	900
aaacataaag	ataaaaaaatt	tacccttggg	aatggggaga	aaaaggaatc		960
tctaccccaa	gaccaaataat	agtaagtcct	gtttctggtc	ctgaacatcc	agaattatgg	1020
aggcttactg	tggcttcaca	ccacattaga	atttggtctg	gaaatcaaac	tttagaaaca	1080
agagatcgta	agccatttta	tactattgac	ctgaattcca	gtctaacagt	tcctttacaa	1140
agttgcgtaa	agccccctta	tatgctagtt	gtaggaaata	tagttattaa	accagactcc	1200
cagactataa	cctgtgaaaa	ttgtagattg	cttacttgca	ttgattcaac	ttttaattgg	1260
caacaccgta	ttctgtgtgt	gagagcaaga	gagggcgtgt	ggatccctgt	gtccatggac	1320
cgaccgtggg	aggcctcgcc	atccgtccat	attttgactg	aagtattaaa	agggtgttta	1380
aatagatcca	aaagattcat	ttttacttta	attgcagtga	ttatgggatt	aattgcagtc	1440
acagctacgg	ctgctgtagc	aggagttgca	ttgcactctt	ctgttcagtc	agtaaacttt	1500
gttaatgatt	ggcaaaaaaa	ttctacaaga	ttgtggaatt	cacaatctag	tattgatcaa	1560
aaattggcaa	atcaaattaa	tgatcttaga	caaactgtca	tttggatggg	agacagactc	1620
atgagcttag	aacatcgttt	ccagttacaa	tgtgactgga	atacgtcaga	tttttgtatt	1680
acaccccaaa	tttataatga	gtctgagcat	cactgggaca	tggttagacg	ccatctacag	1740
ggaagagaag	ataatctcac	tttagacatt	tccaaattaa	aagaacaaat	tttcgaagca	1800
tcaaaagccc	atttaaattt	ggtgccagga	actgaggcaa	ttgcaggagt	tgctgatggc	1860
ctcgcaaatc	ttaaccctgt	cacttggtgt	aagaccattg	gaagtactac	gattataaat	1920
ctcatattaa	tcctgtgtgt	cctgttttgt	ctgttgttag	tctgcagggt	taccaaacag	1980
ctccgaagag	acagcgacca	tcgagaacgg	gccatgatga	cgatggcggt	tttgtcgaaa	2040
agaaaagggg	gaaatgtggg	gaaaagcaag	agagatcaga	ttgttactgt	gtctgtggcc	2100
taa						2103

<210> 82
 <211> 2103
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified env sequence

<400> 82

Substitute Sequence Listing_USSN 10587032_PP019482.007

atgaacccca	gcgagatgca	gcgcaaggcc	cccccccgcc	gccgcccga	ccgcaaccgc	60
gccccctga	cccacaagat	gaacaagatg	gtgaccagcg	aggagcagat	gaagctgccc	120
agcaccaaga	aggccgagcc	ccccacctgg	gcccagctga	agaagctgac	ccagctggcc	180
accaagtacc	tggagaacac	caaggtgacc	cagacccccg	agagcatgct	gctggccgcc	240
ctgatgatcg	tgagcatggt	ggtgagcctg	cccatgcccg	ccggcgccgc	cgccgccaac	300
tacacctact	gggcctacgt	gcccttcccc	cccctgatcc	gcgccgtgac	ctggatggac	360
aaccccaccg	aggtgtacgt	gaacgacagc	gtgtgggtgc	ccggcccat	cgacgaccgc	420
tgccccgcca	agccccagga	ggagggcattg	atgatcaaca	tcagcatcgg	ctaccactac	480
cccccatct	gcctgggccc	cgcccccgcc	tgcctgatgc	ccgcccgtga	gaactggctg	540
gtggaggtgc	ccaccgtgag	ccccatctgc	cgcttcacct	accacatggt	gagcggcatg	600
agcctgcgcc	cccgcgtgaa	ctacctgcag	gacttcagct	accagcgag	cctgaagttc	660
cgccccaaag	gcaagccctg	ccccaaaggag	atccccaaag	agagcaagaa	caccgaggtg	720
ctggtgtggg	aggagtgcgt	ggccaacagc	gccgtgatcc	tgcagaacaa	cgagtccggc	780
accatcatcg	actgggcccc	ccgcggccag	ttctaccaca	actgcagcgg	ccagaccag	840
agctgcccc	gcgcccaggt	gagccccgcc	gtggacagcg	acctgaccga	gagcctggac	900
aagcacaagc	acaagaagct	gcagagcttc	taccctggg	agtggggcga	gaagggcatc	960
agaccccc	gccccaaagat	cgtgagcccc	gtgagcggcc	ccgagcacc	cgagctgtgg	1020
cgctgaccg	tggccagcca	ccacatccgc	atctggagcg	gcaaccagac	cctggagacc	1080
cgcgaccgca	agcccttcta	caccatcgac	ctgaacagca	gcctgaccgt	gcccctgcag	1140
agtgcgtga	agcccccta	catgctgggt	gtgggcaaca	tcgtgatcaa	gcccgcagac	1200
cagaccatca	cctgcgagaa	ctgccgcctg	ctgacctgca	tcgacagcac	cttcaactgg	1260
cagcaccgca	tcctgctggt	gcgcgcccgc	gagggcgtgt	ggatccccgt	gagcatggac	1320
cgccccctgg	aggccagccc	cagcgtgcac	atcctgaccg	aggtgctgaa	gggcgtgctg	1380
aaccgcagca	agcgcttcat	cttcaccctg	atcgccgtga	tcatgggcct	gatcgccgtg	1440
accgccaccg	ccgccgtggc	cggcgtggcc	ctgcacagca	gcgtgcagag	cgtgaacttc	1500
gtgaacgact	ggcagaagaa	cagcaccgcg	ctgtggaaca	gccagagcag	catcgaccag	1560
aagctggcca	accagatcaa	cgacctgcgc	cagaccgtga	tctggatggg	cgaccgcctg	1620
atgagcctgg	agcaccgctt	ccagctgcag	tgcgactgga	acaccagcga	cttctgcatc	1680
accccccaga	tctacaacga	gagcgagcac	cactgggaca	tggatgcgcc	ccacctgcag	1740
ggccgcgagg	acaacctgac	cctggacatc	agcaagctga	aggagcagat	cttcgaggcc	1800
agcaaggccc	acctgaacct	ggtgcccggc	accgaggcca	tcgccggcgt	ggccgacggc	1860
ctggccaacc	tgaaccccg	gacctgggtg	aagaccatcg	gcagcaccac	catcatcaac	1920
ctgatcctga	tcctgggtg	cctgttctgc	ctgctgctgg	tgtgccgctg	caccagcag	1980
ctgcgccgcg	acagcgacca	ccgcgagcgc	gccatgatga	ccatggccgt	gctgagcaag	2040
cgcaagggcg	gcaacgtggg	caagagcaag	cgcgaccaga	tcgtgaccgt	gagcgtggcc	2100
taa						2103

<210>

83

<211>

700

<212>

PRT

<213>

Human endogenous retrovirus, K family (HERV-K)

<400>

83

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65 70 75 80

Leu Met Ile Val Ser Met Val Val Ser Leu Pro Met Pro Ala Gly Ala
85 90 95

Ala Ala Ala Asn Tyr Thr Tyr Trp Ala Tyr Val Pro Phe Pro Pro Leu
100 105 110

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ile Arg Ala Val Thr Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn
115 120 125

Asp Ser Val Trp Val Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys
130 135 140

Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr
145 150 155 160

Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val
165 170 175

Gln Asn Trp Leu Val Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe
180 185 190

Thr Tyr His Met Val Ser Gly Met Ser Leu Arg Pro Arg Val Asn Tyr
195 200 205

Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu Lys Phe Arg Pro Lys Gly
210 215 220

Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu Ser Lys Asn Thr Glu Val
225 230 235 240

Leu Val Trp Glu Glu Cys Val Ala Asn Ser Ala Val Ile Leu Gln Asn
245 250 255

Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala Pro Arg Gly Gln Phe Tyr
260 265 270

His Asn Cys Ser Gly Gln Thr Gln Ser Cys Pro Ser Ala Gln Val Ser
275 280 285

Pro Ala Val Asp Ser Asp Leu Thr Glu Ser Leu Asp Lys His Lys His
290 295 300

Lys Lys Leu Gln Ser Phe Tyr Pro Trp Glu Trp Gly Glu Lys Gly Ile
305 310 315 320

Ser Thr Pro Arg Pro Lys Ile Val Ser Pro Val Ser Gly Pro Glu His
325 330 335

Pro Glu Leu Trp Arg Leu Thr Val Ala Ser His His Ile Arg Ile Trp
340 345 350

Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp Arg Lys Pro Phe Tyr Thr
355 360 365

Ile Asp Leu Asn Ser Ser Leu Thr Val Pro Leu Gln Ser Cys Val Lys
370 375 380

Pro Pro Tyr Met Leu Val Val Gly Asn Ile Val Ile Lys Pro Asp Ser
385 390 395 400

Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu Leu Thr Cys Ile Asp Ser
405 410 415

Thr Phe Asn Trp Gln His Arg Ile Leu Leu Val Arg Ala Arg Glu Gly
420 425 430

Val Trp Ile Pro Val Ser Met Asp Arg Pro Trp Glu Ala Ser Pro Ser
435 440 445

Substitute Sequence Listing_USSN 10587032_PP019482.007

Val His Ile Leu Thr Glu Val Leu Lys Gly Val Leu Asn Arg Ser Lys
 450 455 460
 Arg Phe Ile Phe Thr Leu Ile Ala Val Ile Met Gly Leu Ile Ala Val
 465 470 475 480
 Thr Ala Thr Ala Ala Val Ala Gly Val Ala Leu His Ser Ser Val Gln
 485 490 495
 Ser Val Asn Phe Val Asn Asp Trp Gln Lys Asn Ser Thr Arg Leu Trp
 500 505 510
 Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu Ala Asn Gln Ile Asn Asp
 515 520 525
 Leu Arg Gln Thr Val Ile Trp Met Gly Asp Arg Leu Met Ser Leu Glu
 530 535 540
 His Arg Phe Gln Leu Gln Cys Asp Trp Asn Thr Ser Asp Phe Cys Ile
 545 550 555 560
 Thr Pro Gln Ile Tyr Asn Glu Ser Glu His His Trp Asp Met Val Arg
 565 570 575
 Arg His Leu Gln Gly Arg Glu Asp Asn Leu Thr Leu Asp Ile Ser Lys
 580 585 590
 Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys Ala His Leu Asn Leu Val
 595 600 605
 Pro Gly Thr Glu Ala Ile Ala Gly Val Ala Asp Gly Leu Ala Asn Leu
 610 615 620
 Asn Pro Val Thr Trp Val Lys Thr Ile Gly Ser Thr Thr Ile Ile Asn
 625 630 635 640
 Leu Ile Leu Ile Leu Val Cys Leu Phe Cys Leu Leu Leu Val Cys Arg
 645 650 655
 Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp His Arg Glu Arg Ala Met
 660 665 670
 Met Thr Met Ala Val Leu Ser Lys Arg Lys Gly Gly Asn Val Gly Lys
 675 680 685
 Ser Lys Arg Asp Gln Ile Val Thr Val Ser Val Ala
 690 695 700